

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:59:34 ; Search time 87.3 seconds  
(without alignments)  
5.671 Million cell updates/sec

Title: US-09-290-049A-19

Perfect score: 110

Sequence: 1 VPSYSFARAHSEVQDIIRDII 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Number of hits satisfying chosen parameters: 212252

Maximum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA.\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*

2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*

3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*

4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*

5: /cgn2\_6/ptodata/2/iaa/PTUS\_COMB.pep.\*

6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	98	89.1	1375	4	US-09-210-361-4
2	97	88.2	1475	3	US-09-007-999-2
3	97	88.2	1475	4	US-09-210-361-2
4	64	58.2	1430	3	US-09-008-172-2
5	64	58.2	1430	4	US-09-210-361-6
6	64	58.2	1577	2	US-08-793-824-2
7	45.5	41.4	484	4	US-08-913-578-2
8	45.5	41.4	484	4	US-08-785-427-2
9	43	39.1	2627	2	US-08-751-189-3
10	43	39.1	2627	2	US-09-060-836-3
11	43	39.1	2627	4	US-09-184-445-3
12	42	38.2	196	2	US-08-684-024-2
13	42	38.2	196	2	US-08-684-024-9
14	42	38.2	196	3	US-09-145-868-2
15	42	38.2	196	3	US-09-145-868-9
16	41	37.3	639	2	US-08-557-309B-37
17	41	37.3	639	3	US-08-834-306-37
18	41	37.3	639	4	US-08-993-674A-37
19	40.5	36.8	348	2	US-08-844-153-2
20	40.5	36.8	348	2	US-08-962-203-2
21	40.5	36.8	480	4	US-09-282-125A-2
22	40.5	36.8	480	4	US-09-273-142-2
23	40	36.4	1068	3	US-08-390-874C-11
24	40	36.4	1068	4	US-09-265-772-11
25	40	36.4	1069	2	US-08-162-081B-37
26	40	36.4	1069	2	US-08-780-872-37
27	40	36.4	1069	4	US-09-085-957-37

28	40	36.4	1080	2	US-08-162-081B-36
29	40	36.4	1080	2	US-08-780-872-36
30	40	36.4	1080	4	US-09-085-957-36
31	39	35.5	184	1	US-08-353-550-2
32	39	35.5	184	2	US-08-551-687-2
33	39	35.5	236	3	US-08-935-263-2
34	39	35.5	428	1	US-08-353-550-1
35	39	35.5	428	2	US-08-551-687-1
36	39	35.5	431	3	US-08-807-342B-5
37	39	35.5	469	1	US-08-353-550-6
38	39	35.5	469	2	US-08-551-687-6
39	39	35.5	933	1	US-08-370-193A-8
40	39	35.5	1365	6	5194600-4
41	38	34.5	43	4	US-08-679-006-2
42	38	34.5	348	4	US-09-198-955A-8
43	38	34.5	418	5	PCT-US94-01321-72
44	38	34.5	1454	4	US-08-392-459-22
45	38	34.5	1454	4	US-08-392-459-26

ALIGNMENTS

RESULT 1

US-09-210-361-4

; Sequence 4, Application US/09210361

; Patent No. 6284479

; GENERAL INFORMATION:

; APPLICANT: Nichols, Scott E.

; TITLE OF INVENTION: Substitutes for Modified Starches and

; FILE OF INVENTION: Latexes in Paper Manufacture

; FILE REFERENCE: 0357CR

; CURRENT APPLICATION NUMBER: US/09/210.361

; EARLIER FILING DATE: 1998-12-11

; EARLIER APPLICATION NUMBER: 09/007,999

; EARLIER FILING DATE: 1998-01-16

; EARLIER APPLICATION NUMBER: 08/478,704

; EARLIER FILING DATE: 1995-06-07

; EARLIER APPLICATION NUMBER: 09/009,620

; EARLIER FILING DATE: 1998-01-20

; EARLIER APPLICATION NUMBER: 08/485,243

; EARLIER FILING DATE: 1995-06-07

; EARLIER APPLICATION NUMBER: 09/008,172

; EARLIER FILING DATE: 1998-01-16

; EARLIER APPLICATION NUMBER: 08/482,711

; EARLIER FILING DATE: 1995-06-07

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 4

; LENGTH: 137

; TYPE: PRT

; ORGANISM: str:ptococcus mutans

US-09-210-361-4

Query Match 89.1%; Score 98; DB 4; Length 1375;

Best Local Similarity 86.4%; Pred. No. 4.4e-08;

Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHSEVQDIIRDII 22

Db 578 VPSYSFARAHSEVQDIIRDII 599

RESULT 2

US-09-007-999-2

; Sequence 2, Application US/09007999

; Patent No. 6087559

; GENERAL INFORMATION:

; APPLICANT: Nichols, Scott E.

; TITLE OF INVENTION: Substitutes for Modified Starch and

; FILE OF INVENTION: Latexes in Paper Manufacture

; FILE REFERENCE: 0356D

; CURRENT APPLICATION NUMBER: US/09/007,999  
; CURRENT FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-007-999-2

Query Match 88.2%; Score 97; DB 3; Length 1475;  
Best Local Similarity 86.4%; Pred. No. 7.1e-08;  
Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 VPSYFARHDSVQDIIRDII 22  
||||| |||||||: |||  
Db 552 VPSYFIRAHDSVQDLIADII 573

RESULT 3  
US-09-210-361-2  
; Sequence 2, Application US/09210361  
; Patent No. 6284479  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starches and  
; FILE REFERENCE: 0357CR  
; CURRENT APPLICATION NUMBER: US/09/210,361  
; CURRENT FILING DATE: 1998-12-11  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/009,620  
; EARLIER FILING DATE: 1998-01-20  
; EARLIER APPLICATION NUMBER: 08/485,243  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/008,172  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-210-361-2

Query Match 88.2%; Score 97; DB 4; Length 1475;  
Best Local Similarity 86.4%; Pred. No. 7.1e-08;  
Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 VPSYFARHDSVQDIIRDII 22  
||||| |||||||: |||  
Db 552 VPSYFIRAHDSVQDLIADII 573

RESULT 4  
US-09-008-172-2  
; Sequence 2, Application US/09008172  
; Patent No. 6127602  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; FILE REFERENCE: 0358D  
; CURRENT APPLICATION NUMBER: US/09/008,172

; CURRENT FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-008-172-2

Query Match 58.2%; Score 64; DB 3; Length 1430;  
Best Local Similarity 65.0%; Pred. No. 0.023;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 3 SYSFARHDSVQDIIRDII 22  
:| | |||||: | |  
Db 576 NYIFRAHDSVQTVIAKII 595

RESULT 5  
US-09-210-361-6  
; Sequence 6, Application US/09210361  
; Patent No. 6284479  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starches and  
; FILE REFERENCE: 0357CR  
; CURRENT APPLICATION NUMBER: US/09/210,361  
; CURRENT FILING DATE: 1998-12-11  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/009,620  
; EARLIER FILING DATE: 1998-01-20  
; EARLIER APPLICATION NUMBER: 08/485,243  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/008,172  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 6  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-210-361-6

Query Match 58.2%; Score 64; DB 4; Length 1430;  
Best Local Similarity 65.0%; Pred. No. 0.023;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 3 SYSFARHDSVQDIIRDII 22  
:| | |||||: | |  
Db 576 NYIFRAHDSVQTVIAKII 595

RESULT 6  
US-08-793-824-2  
; Sequence 2, Application US/08793824  
; Patent No. 5981838  
; GENERAL INFORMATION:  
; APPLICANT: Simpson, Christine Lynn  
; APPLICANT: Giffard, Phillip Morrison  
; APPLICANT: Jacques, Nicholas Anthony  
; TITLE OF INVENTION: Genetic Manipulation of plants to  
; FILE REFERENCE: 0358D  
; TITLE OF INVENTION: Increase Stored Carbohydrates  
; NUMBER OF SEQUENCES: 2

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Griffith Hack & Co  
;; STREET: Level 8, 168 Walker Street  
;; CITY: No. 5981838th Sydney  
;; STATE: New South Wales  
;; COUNTRY: Australia  
;; ZIP: 2060  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/793.824  
;; FILING DATE:  
;; CLASSIFICATION: 800  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: AU PM7643  
;; FILING DATE: 24-AUG-1994  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 61 2 9957 5944  
;; TELEFAX: 61 2 957 6288  
;; TELEX: 26547  
;;  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1577 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: not relevant  
;; MOLECULE TYPE: protein  
;; ORIGINAL SOURCE:  
;; ORGANISM: Streptococcus salivarius  
;;  
US-08-793-824-2

Query Match 58.2%; Score 64; DB 2; Length 1577;  
Best Local Similarity 60.0%; Pred. No. 0.026;  
Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 3 SYSFARHDSVQDIIRDII 22  
Db 661 NYIFVRHDSVQAVLANII 680

RESULT 7  
US-08-913-578-2  
; Sequence 2, Application US/08913578  
; Patent No. 6218159  
; GENERAL INFORMATION:  
; APPLICANT: Hodgson, John  
; APPLICANT: Lawlor, Elizabeth  
; TITLE OF INVENTION: No. 6218159el tRNA synthetase  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406-0939  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/913.578  
; FILING DATE: 17-JAN-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 9601069.9  
; FILING DATE: 19-JAN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gimmi, Edward R  
; REGISTRATION NUMBER: 38,891  
; REFERENCE/DOCKET NUMBER: P31352  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-270-4478  
; TELEFAX: 610-270-5090  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 484 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-913-578-2

Query Match 41.4%; Score 45.5; DB 4; Length 484;

;; NAME: Gimmi, Edward R  
;; REGISTRATION NUMBER: 38,891  
;; REFERENCE/DOCKET NUMBER: P31352  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 610-270-4478  
;; TELEFAX: 610-270-5090  
;; TELEX:  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 484 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; US-08-913-578-2

Query Match 41.4%; Score 45.5; DB 4; Length 484;  
Best Local Similarity 40.9%; Pred. No. 8.1;  
Matches 9; Conservative 6; Mismatches 4; Indels 3; Gaps 1;

Qy 1 VPSYSFARA---HDSVQDIIR 19  
Db 193 IPTYNFAVAIDHDHYNQISDVIR 214

RESULT 8  
US-08-785-427-2  
; Sequence 2, Application US/08785427  
; Patent No. 6238900  
; GENERAL INFORMATION:  
; APPLICANT: Hodgson, John  
; APPLICANT: Lawlor, Elizabeth  
; TITLE OF INVENTION: No. 6238900el tRNA synthetase  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406-0939  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/785.427  
; FILING DATE: 17-JAN-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 9601069.9  
; FILING DATE: 19-JAN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gimmi, Edward R  
; REGISTRATION NUMBER: 38,891  
; REFERENCE/DOCKET NUMBER: P31352  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-270-4478  
; TELEFAX: 610-270-5090  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 484 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-785-427-2

Query Match 41.4%; Score 45.5; DB 4; Length 484;

Best Local Similarity 40.9%; Pred. No. 8.1;  
Matches 9; Conservative 6; Mismatches 4; Indels 3; Gaps 1;

QY 1 VPSYSFARA---HSEVQDIIR 19  
:|:|:|:| | | :|:|:|  
Db 193 IPTYNFAVAIDDDHYMQISDVIR 214

RESULT 9  
US-08-751-189-3  
; Sequence 3, Application US/08/751189  
; Patent No. 5919656  
; GENERAL INFORMATION:  
; APPLICANT: Harrington, Lea A.  
; APPLICANT: Robinson, Murray O.  
; TITLE OF INVENTION: No. 5919656el Genes Encoding Telomerase Protein  
; NUMBER OF SEQUENCES: 1  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen, Inc.  
; STREET: 1840 De Havilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/751,189  
; FILING DATE: 15-NOV-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oleski, Nancy A.  
; REGISTRATION NUMBER: 34,688  
; REFERENCE/DOCKET NUMBER: A-433  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2627 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-751-189-3

Query Match 39.1%; Score 43; DB 2; Length 2627;  
Best Local Similarity 41.2%; Pred. No. 1.6e+02;  
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 VPSYSFARAHDSEVQDI 17  
:||||: :|:|:  
Db 205 MPSYSLSGLEEEVEDL 221

RESULT 10  
US-09-060-836-3  
; Sequence 3, Application US/09060836  
; Patent No. 5981707  
; GENERAL INFORMATION:  
; APPLICANT: Harrington, Lea A.  
; APPLICANT: Robinson, Murray O.  
; TITLE OF INVENTION: No. 5981707el Genes Encoding Telomerase Protein  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen, Inc.  
; STREET: 1840 De Havilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA

ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/060,836  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/751,189  
; FILING DATE: 15-NOV-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oleski, Nancy A.  
; REGISTRATION NUMBER: 34,688  
; REFERENCE/DOCKET NUMBER: A-433  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2627 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-09-060-836-3

Query Match 39.1%; Score 43; DB 2; Length 2627;  
Best Local Similarity 41.2%; Pred. No. 1.6e+02;  
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 VPSYSFARAHDSEVQDI 17  
:||||: :|:|:  
Db 205 MPSYSLSGLEEEVEDL 221

RESULT 11  
US-09-184-445-3  
; Sequence 3, Application US/09184445  
; Patent No. 6174703  
; GENERAL INFORMATION:  
; APPLICANT: Harrington, Lea A.  
; APPLICANT: Robinson, Murray O.  
; TITLE OF INVENTION: No. 6174703el Genes Encoding Telomerase Protein  
; TITLE OF INVENTION: 1  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen, Inc.  
; STREET: 1840 De Havilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/184,445  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/751,189  
; FILING DATE: 15-NOV-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oleski, Nancy A.  
; REGISTRATION NUMBER: 34,688  
; REFERENCE/DOCKET NUMBER: A-433  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2627 amino acids  
; TYPE: amino acid



STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-09-184-445-3

Query Match 39.1%; Score 43; DB 4; Length 2627;  
Best Local Similarity 41.2%; Pred. No. 1.6e+02;  
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 VPSYFARAHDEVODI 17  
:||||: :|||:  
DB 205 MPYSLSLGEVEEDL 221

## RESULT 12

US-08-684-024-2  
Sequence 2, Application US/08684024  
Patent No. 5834298  
GENERAL INFORMATION:

APPLICANT: Beneza, Robert  
TITLE OF INVENTION: GENE ENCODING THE HUMAN HOMOLOG OF MAD2  
NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/684.024

FILING DATE: 19-JUL-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 1747/46621-A  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 278-0400

TELEFAX: (212) 391-0526

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 196 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-684-024-2

Query Match 38.2%; Score 42; DB 2; Length 196;  
Best Local Similarity 46.7%; Pred. No. 11;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 8 RAHDEVQDIIRDII 22  
:|||:|:|:  
DB 49 KTHDDLKDYIRKIL 63

## RESULT 13

US-08-684-024-9

Sequence 9, Application US/08684024

Patent No. 5834298

GENERAL INFORMATION:

APPLICANT: Beneza, Robert

TITLE OF INVENTION: GENE ENCODING THE HUMAN HOMOLOG OF MAD2

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/684.024

FILING DATE: 19-JUL-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 1747/46621-A

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 278-0400

TELEFAX: (212) 391-0526

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 196 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-684-024-9

Query Match 38.2%; Score 42; DB 2; Length 196;  
Best Local Similarity 46.7%; Pred. No. 11;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 8 RAHDEVQDIIRDII 22  
:|||:|:|:  
DB 49 KTHDDLKDYIRKIL 63

## RESULT 14

US-09-145-868-2

Sequence 2, Application US/09145868

Patent No. 6096522

GENERAL INFORMATION:

APPLICANT: Beneza, Robert

TITLE OF INVENTION: GENE ENCODING THE HUMAN HOMOLOG OF MAD2

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham LLP

STREET: 1185 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/145.868

FILING DATE: 02-SEP-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 1747/46621-B

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 278-0400

TELEFAX: (212) 391-0526

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 196 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-145-868-2

Query Match 38.2%; Score 42; DB 3; Length 196;  
Best Local Similarity 46.7%; Pred. No. 11;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 8 RAHDSEVQDIIRDII 22  
DB 49 KTHDDELKDYIRKIL 63

## ULT 15

US-09-145-868-9  
Sequence 9, Application US/09145868  
Patent No. 6096522  
GENERAL INFORMATION:  
APPLICANT: Benezira, Robert  
TITLE OF INVENTION: GENE ENCODING THE HUMAN HOMOLOG OF MAD2  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/145,868  
FILING DATE: 02-SEP-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 1747/46621-B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0526  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 196 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-145-868-9

Query Match 38.2%; Score 42; DB 3; Length 196;  
Best Local Similarity 46.7%; Pred. No. 11;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 8 RAHDSEVQDIIRDII 22  
DB 49 KTHDDELKDYIRKIL 63

Search completed: March 27, 2002, 13:59:35  
Job time: 588 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:20:30 ; Search time 1139.61 seconds  
(without alignments)  
5.360 Million cell updates/sec

Title: US-09-290-049A-19

Perfect score: 110

Sequence: 1 VPSYSFARAHSEVDIIRDI 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 3148936 seqs, 277657034 residues

Number of hits satisfying chosen parameters: 3148936

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Pending\_Patents\_AA\_Main:\*
- 1: /cgn2\_6/ptodata/2/paa/PCTUS\_COMB.pap.\*
  - 2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pap.\*
  - 3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pap.\*
  - 4: /cgn2\_6/ptodata/2/paa/US08\_COMB.pap.\*
  - 5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pap.\*
  - 6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pap.\*
  - 7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pap.\*
  - 8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pap.\*
  - 9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pap.\*
  - 10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pap.\*
  - 11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pap.\*
  - 12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pap.\*
  - 13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pap.\*
  - 14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pap.\*
  - 15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pap.\*
  - 16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pap.\*
  - 17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pap.\*
  - 18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pap.\*
  - 19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pap.\*
  - 20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pap.\*
  - 21: /cgn2\_6/ptodata/2/paa/US097\_COMB.pap.\*
  - 22: /cgn2\_6/ptodata/2/paa/US098\_COMB.pap.\*
  - 23: /cgn2\_6/ptodata/2/paa/US099\_COMB.pap.\*
  - 24: /cgn2\_6/ptodata/2/paa/US60\_COMB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	110	100.0	22	16	US-09-290-049-19
2	108	98.2	22	16	US-09-290-049-17
3	98	89.1	22	16	US-09-290-049-15
4	98	89.1	1375	21	US-09-740-274-4
5	97	88.2	1475	19	US-09-557-848-2
6	97	88.2	1475	21	US-09-740-274-2
7	73.5	66.8	20	16	US-09-290-049-2
8	69	62.7	22	16	US-09-290-049-18
9	68	61.8	2057	18	US-09-499-203-2

10	64	58.2	22	16	US-09-290-049-16	Sequence 16, Appli
11	64	58.2	1430	20	US-09-649-885-2	Sequence 2, Appli
12	64	58.2	1430	21	US-09-740-274-6	Sequence 6, Appli
13	47.5	43.2	481	24	US-60-242-578-904	Sequence 904, App
14	47.5	43.2	481	24	US-60-253-625-2248	Sequence 2248, Ap
15	47.5	43.2	481	24	US-60-257-931-3142	Sequence 3142, Ap
16	47.5	43.2	481	24	US-60-269-308-4164	Sequence 4164, Ap
17	47	42.7	194	18	US-09-417-507-43716	Sequence 43716, A
18	47	42.7	214	1	PCT-US01-08631-31600	Sequence 31600, A
19	46.5	42.3	501	15	US-09-134-001C-4115	Sequence 4115, Ap
20	46.5	42.3	501	18	US-09-450-969-5378	Sequence 5378, Ap
21	46	41.8	83	24	US-60-207-215-374	Sequence 374, App
22	46	41.8	94	24	US-60-248-798-269	Sequence 269, App
23	46	41.8	129	16	US-09-270-767-32878	Sequence 32878, A
24	46	41.8	129	16	US-09-270-767-48095	Sequence 48095, A
25	46	41.8	129	16	US-09-270-849B-188016	Sequence 188016,
26	46	41.8	129	24	US-60-196-710-6423	Sequence 6423, Ap
27	46	41.8	164	22	US-09-834-366-17052	Sequence 17052, A
28	46	41.8	164	24	US-60-197-873-17052	Sequence 17052, A
29	46	41.8	188	21	US-09-758-460-507	Sequence 507, App
30	46	41.8	208	24	US-60-196-713-3972	Sequence 3972, Ap
31	46	41.8	309	17	US-09-345-473B-37	Sequence 37, Appl
32	46	41.8	309	17	US-09-345-473C-37	Sequence 37, Appl
33	46	41.8	309	20	US-09-609-360B-37	Sequence 37, Appl
34	46	41.8	309	22	US-09-862-027-37	Sequence 37, Appl
35	46	41.8	479	1	PCT-US01-08631-51319	Sequence 51319, A
36	46	41.8	480	7	US-08-350-584-58	Sequence 58, Appl
37	46	41.8	490	1	PCT-US01-03800A-2103	Sequence 2103, Ap
38	46	41.8	586	23	US-09-948-933-369	Sequence 369, App
39	46	41.8	597	24	US-60-213-847-886	Sequence 886, App
40	46	41.8	724	1	PCT-US01-04098A-1860	Sequence 1860, Ap
41	46	41.8	743	24	US-60-202-090-1	Sequence 1, Appli
42	46	41.8	788	24	US-60-230-445-1292	Sequence 1292, Ap
43	45.5	41.4	484	11	US-08-796-627-2	Sequence 2, Appli
44	45.5	41.4	484	13	US-08-937-906-2	Sequence 2, Appli
45	45.5	41.4	484	14	US-09-010-962-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-09-290-049-19  
; Sequence 19, Application US/09290049  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CARRIES  
; FILE REFERENCE: PDC98-01p2A  
; CURRENT APPLICATION NUMBER: US/09/290,049  
; CURRENT FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081,550  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/115,142  
; EARLIER FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 19  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. sobrinus  
US-09-290-049-19

Query Match	100.0%	Score 110;	DB 16;	Length 22;
Best Local Similarity	100.0%	Pred. No. 3	6e-11;	
Matches	22;	Conservative	0;	Mismatches 0;
		Indels	0;	Gaps 0;
Qy	1	VPSYSFARAHSEVDIIRDI	22	
Db	1	VPSYSFARAHSEVDIIRDI	22	

RESULT 2  
US-09-290-049-17  
; Sequence 17, Application US/09290049  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE OF INVENTION: CARIES  
; FILE REFERENCE: FDC98-012A  
; CURRENT APPLICATION NUMBER: US/09/290,049  
; CURRENT FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081,550  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/115,142  
; EARLIER FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 17  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. downei  
US-09-290-049-17

Query Match 98.1%; Score 108; DB 16; Length 22;  
Best Local Similarity 95.5%; Pred. No. 7.9e-11;  
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VPSYSFARAHSEVQDIIRDII 22  
||||| |||||||||:|||||  
DB 1 VPSYSFARAHSEVQDLIRNII 22

RESULT 3  
US-09-290-049-15  
; Sequence 15, Application US/09290049  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE OF INVENTION: CARIES  
; FILE REFERENCE: FDC98-012A  
; CURRENT APPLICATION NUMBER: US/09/290,049  
; CURRENT FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081,550  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/115,142  
; EARLIER FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 15  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. mutans  
US-09-290-049-15

Query Match 89.1%; Score 98; DB 16; Length 22;  
Best Local Similarity 86.4%; Pred. No. 3.9e-09;  
Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 VPSYSFARAHSEVQDIIRDII 22  
||||| |||||||||:|||||  
DB 1 VPSYSFARAHSEVQDLIRNII 22

RESULT 4  
US-09-740-274-4  
; Sequence 4, Application US/09740274  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper

; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 1375  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-740-274-4

Query Match 89.1%; Score 98; DB 21; Length 1375;  
Best Local Similarity 86.4%; Pred. No. 8.6e-07;  
Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 VPSYSFARAHSEVQDIIRDII 22  
||||| |||||||||:|||||  
DB 578 VPSYSFARAHSEVQDLIRNII 599

RESULT 5  
US-09-557-848-2  
; Sequence 2, Application US/09557848  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; FILE OF INVENTION: Latexes in Paper Manufacture  
; FILE REFERENCE: 0356D2  
; CURRENT APPLICATION NUMBER: US/09/557,848  
; CURRENT FILING DATE: 2000-04-26  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-557-848-2

Query Match 88.2%; Score 97; DB 19; Length 1475;  
Best Local Similarity 86.4%; Pred. No. 1.4e-06;  
Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 VPSYSFARAHSEVQDIIRDII 22  
||||| |||||||||:|||||  
DB 552 VPSYSFARAHSEVQDLIRNII 573

RESULT 6  
US-09-740-274-2  
; Sequence 2, Application US/09740274  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper

```

; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2

```

```

Query Match      88.2%; Score 97; DB 21; Length 1475;
Best Local Similarity 86.4%; Pred. No. 1.4e-06;
Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 VPSYSFARAHSEVQDIIRDII 22
    ||||| ||||| ||||| |||||
Db 552 VPSYSFIRAHSEVQDLIADII 573

```

```

RESULT 7
US-09-290-049-2
; Sequence 2, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CARRIES
; FILE REFERENCE: FDC98-01p2A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: HDS peptide
US-09-290-049-2

```

```

Query Match      66.8%; Score 73.5; DB 16; Length 20;
Best Local Similarity 84.2%; Pred. No. 4.9e-05;
Matches 16; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

```

```

QY 1 VPSYSFAR-AHSEVQDII 18
    ||||| ||||| ||||| |||||
Db 1 VPSYSFIRAHSEVQDLI 19

```

```

RESULT 8
US-09-290-049-18
; Sequence 18, Application US/09290049

```

```

; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CARRIES
; FILE REFERENCE: FDC98-01p2A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 22
; TYPE: PRT
; ORGANISM: S. downei
US-09-290-049-18

```

```

Query Match      62.7%; Score 69; DB 16; Length 22;
Best Local Similarity 68.2%; Pred. No. 0.00032;
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

```

```

QY 1 VPSYSFARAHSEVQDIIRDII 22
    ||||| ||||| ||||| |||||
Db 1 VPSYSFIRAHSEVQTRIADII 22

```

```

RESULT 9
US-09-499-203-2
; Sequence 2, Application US/09499203
; GENERAL INFORMATION:
; APPLICANT: KOSSHANN, Jens
; APPLICANT: WELSH, Thomas
; APPLICANT: QUANZ, Martin
; APPLICANT: KNUTH, Karola
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase
; FILE REFERENCE: 147-196P
; CURRENT APPLICATION NUMBER: US/09/499,203
; CURRENT FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2057
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-499-203-2

```

```

Query Match      61.8%; Score 68; DB 18; Length 2057;
Best Local Similarity 63.2%; Pred. No. 0.18;
Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 1 VPSYSFARAHSEVQDIIR 19
    ||||| ||||| ||||| |||||
Db 757 IPNYSFVRAHDYADQPIR 775

```

```

RESULT 10
US-09-290-049-16
; Sequence 16, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CARRIES
; FILE REFERENCE: FDC98-01p2A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13

```

; EARLIER APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 16  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. mutans  
US-09-290-049-16

Query Match 58.2%; Score 64; DB 16; Length 22;  
Best Local Similarity 65.0%; Pred. No. 0.0023;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 3 SYSFARHDSVQDIIRII 22  
:| | | | | | | | | | | | | | | | | | | |  
Db 3 NYIFIRHDSVQTVIAKII 22

RESULT 11  
US-09-649-885-2  
; Sequence 2, Application US/09649885  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; FILE REFERENCE: 035822  
; CURRENT APPLICATION NUMBER: US/09/649,885  
; CURRENT FILING DATE: 2000-08-28  
; PRIOR APPLICATION NUMBER: US 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-649-885-2

Query Match 58.2%; Score 64; DB 20; Length 1430;  
Best Local Similarity 65.0%; Pred. No. 0.52;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 3 SYSFARHDSVQDIIRII 22  
:| | | | | | | | | | | | | | | | | | | |  
Db 576 NYIFIRHDSVQTVIAKII 595

RESULT 12  
US-09-740-274-6  
; Sequence 6, Application US/09740274  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16

; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 6  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-740-274-6

Query Match 58.2%; Score 64; DB 21; Length 1430;  
Best Local Similarity 65.0%; Pred. No. 0.52;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 3 SYSFARHDSVQDIIRII 22  
:| | | | | | | | | | | | | | | | | | | |  
Db 576 NYIFIRHDSVQTVIAKII 595

RESULT 13  
US-60-242-578-904  
; Sequence 904, Application US/60242578  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, R  
; APPLICANT: Ohlsen, K. L.  
; APPLICANT: Zyskind, J. W.  
; TITLE OF INVENTION: Genes Identified as essential in  
; FILE REFERENCE: ELITRA.017PR2  
; CURRENT APPLICATION NUMBER: US/60/242,578  
; CURRENT FILING DATE: 2000-10-23  
; NUMBER OF SEQ ID NOS: 1057  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 904  
; LENGTH: 481  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-60-242-578-904

Query Match 43.2%; Score 47.5; DB 24; Length 481;  
Best Local Similarity 45.5%; Pred. No. 79;  
Matches 10; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

Qy 1 VPSYSFARHDS---EVQDIIR 19  
:| | | | | | | | | | | | | | | | | | | |  
Db 193 IPTYNFAVAIDYYMEISDVIR 214

RESULT 14  
US-60-253-625-2248  
; Sequence 2248, Application US/60253625  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, RH  
; APPLICANT: Ohlsen, KL  
; APPLICANT: Zyskind, JW  
; APPLICANT: Trawick, JD  
; APPLICANT: Wall, D  
; TITLE OF INVENTION: Identification of essential genes in  
; TITLE OF INVENTION: Staphylococcus aureus, pseudomonas aeruginosa, Klebsiella  
; FILE REFERENCE: ELITRA.017PR3  
; CURRENT APPLICATION NUMBER: US/60/253,625  
; CURRENT FILING DATE: 2000-11-27  
; NUMBER OF SEQ ID NOS: 2768  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2248  
; LENGTH: 481  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-60-253-625-2248

Query Match 43.2%; Score 47.5; DB 24; Length 481;  
Best Local Similarity 45.5%; Pred. No. 79;  
Matches 10; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

QY 1 VPSYSFARAHDS---EVQDIIR 19  
:|:|:| | | | | | | | | |  
Db 193 IPTYNFAVAIDDDYYMEISDVIR 214

## RESULT 15

US-60-257-931-3142  
; Sequence 3142, Application US/60257931  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, R  
; APPLICANT: Ohlsen, KL  
; APPLICANT: Zyskind, JW  
; APPLICANT: Trawick, JD  
; APPLICANT: Wall, D  
; TITLE OF INVENTION: Genes identified as essential in Staphylococcus aureus, Salmonella typhimurium, Klebsiella pneumoniae and Pseudomonas aeruginosa  
; FILE REFERENCE: ELITRA: 017PR4  
; CURRENT APPLICATION NUMBER: US/60/257,931  
; CURRENT FILING DATE: 2000-12-22  
; NUMBER OF SEQ ID NOS: 3592  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3142  
; LENGTH: 481  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-60-257-931-3142

Query Match 43.2%; Score 47.5; DB 24; Length 481;  
Best Local Similarity 45.5%; Pred. No. 79;  
Matches 10; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

QY 1 VPSYSFARAHDS---EVQDIIR 19  
:|:|:| | | | | | | | | |  
Db 193 IPTYNFAVAIDDDYYMEISDVIR 214

Search completed: March 27, 2002, 14:20:31  
Job time: 1578 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:22:49 ; Search time 137.48 Seconds  
(without alignments)  
11.042 Million cell updates/sec

Title: us-09-290-049A-19  
Perfect score: 110  
Sequence: 1 VPSYSFARAHSEVDIIRDI 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 332938 seqs, 6899538 residues

Total number of hits satisfying chosen parameters: 332938

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_New.\*  
1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep.\*  
8: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	110	100.0	22	6	US-09-562-328-32
2	110	100.0	22	6	US-09-290-049A-19
3	108	98.2	22	6	US-09-562-328-30
4	108	98.2	22	6	US-09-290-049A-17
5	98	89.1	22	6	US-09-562-328-28
6	98	89.1	22	6	US-09-290-049A-15
7	84	76.4	19	6	US-09-562-328-27
8	84	76.4	19	6	US-09-290-049A-2
9	84	76.4	23	6	US-09-562-328-44
10	81	73.6	523	6	US-09-604-957-5
11	73.5	66.8	20	6	US-09-562-328-26
12	69	62.7	22	6	US-09-562-328-31
13	69	62.7	22	6	US-09-290-049A-18
14	68	61.8	535	6	US-09-604-957-7
15	68	61.8	1278	6	US-09-604-957-3
16	64	58.2	22	6	US-09-562-328-29
17	64	58.2	22	6	US-09-290-049A-16
18	64	58.2	545	6	US-09-604-957-4
19	63	57.3	584	6	US-09-604-957-6
20	51	46.4	215	6	US-09-675-784A-8997
21	47.5	43.2	481	1	PCT-US02-03987-5584
22	47.5	43.2	481	6	US-09-815-242-5584
23	47.5	43.2	481	7	US-10-072-851-5584
24	47.5	43.2	487	1	PCT-US02-03987-12456
25	47.5	43.2	487	6	US-09-815-242-12456

26	47.5	43.2	487	7	US-10-072-851-12456
27	46	41.8	309	6	US-09-609-360C-37
28	46	41.8	309	6	US-09-345-473E-37
29	46	41.8	521	6	US-09-646-673A-117
30	46	41.8	564	6	US-09-646-673A-180
31	46	41.8	2193	8	US-60-338-690-5
32	44.5	40.5	486	1	PCT-US02-03987-13455
33	44.5	40.5	486	6	US-09-815-242-13455
34	44.5	40.5	486	7	US-10-072-851-13455
35	43	39.1	274	6	US-09-675-784A-9123
36	43	39.1	430	6	US-09-708-427-28756
37	43	39.1	531	6	US-09-708-427-28755
38	43	39.1	557	5	US-09-506-720B-198
39	43	39.1	571	6	US-09-708-427-28754
40	43	39.1	571	8	US-60-356-051-2543
41	42.5	38.6	183	6	US-09-675-784A-13501
42	42	38.2	102	6	US-09-675-784A-8142
43	42	38.2	219	7	US-10-015-127-10718
44	42	38.2	304	6	US-09-708-427-7240
45	42	38.2	323	6	US-09-708-427-7239

## ALIGNMENTS

### RESULT 1

US-09-562-328-32  
; Sequence 32, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 32  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-32

Query Match Best Local Similarity 100.0%; Score 110; DB 6; Length 22;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 VPSYSFARAHSEVDIIRDI 22

Db 1 VPSYSFARAHSEVDIIRDI 22

### RESULT 2

US-09-290-049A-19  
; Sequence 19, Application US/09290049A  
; GENERAL INFORMATION:  
; APPLICANT: SMITH, DANIEL J.  
; APPLICANT: TAUBMAN, MARTIN A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CARIES  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 19  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. sobrinus  
US-09-290-049a-19

Query Match 100.0%; Score 110; DB 6; Length 22;  
Best Local Similarity 100.0%; Pred. No. 4.9e-11;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHDSVQDIIRDII 22  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 VPSYSFARAHDSVQDIIRDII 22

RESULT 3  
US-09-562-328-30  
; Sequence 30, Application US/09562328  
; GENERAL INFORMATION:

; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 30  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-30

Query Match 98.2%; Score 108; DB 6; Length 22;  
Best Local Similarity 95.5%; Pred. No. 1e-10;  
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHDSVQDIIRDII 22  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 VPSYSFARAHDSVQDIIRDII 22

RESULT 4  
US-290-049a-17  
; Sequence 17, Application US/09290049A  
; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CARRIES  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. downei  
US-09-290-049a-17

Query Match 98.2%; Score 108; DB 6; Length 22;  
Best Local Similarity 95.5%; Pred. No. 1e-10;

Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 VPSYSFARAHDSVQDIIRDII 22  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 VPSYSFARAHDSVQDIIRDII 22

RESULT 5  
US-09-562-328-28  
; Sequence 28, Application US/09562328  
; GENERAL INFORMATION:

; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 28  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-28

Query Match 89.1%; Score 98; DB 6; Length 22;  
Best Local Similarity 86.4%; Pred. No. 3.9e-09;  
Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHDSVQDIIRDII 22  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 VPSYSFARAHDSVQDIIRDII 22

RESULT 6  
US-09-290-049A-15  
; Sequence 15, Application US/09290049A  
; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CARRIES  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. mutans  
US-09-290-049A-15

Query Match 89.1%; Score 98; DB 6; Length 22;  
Best Local Similarity 86.4%; Pred. No. 3.9e-09;  
Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHDSVQDIIRDII 22  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 VPSYSFARAHDSVQDIIRDII 22

RESULT 7  
US-09-562-328-27

; Sequence 27, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 27  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-27

Query Match 76.4%; Score 84; DB 6; Length 19;  
Best Local Similarity 88.9%; Pred. No. 5.3e-07;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSFARHDSEVDII 18  
Db 1 VPSYSFIRHDSEVDLI 18  
||||| |||||||

RESULT 8  
US-09-290-049A-2  
; Sequence 2, Application US/09290049A  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: HDS peptide  
US-09-290-049A-2

Query Match 76.4%; Score 84; DB 6; Length 19;  
Best Local Similarity 88.9%; Pred. No. 5.3e-07;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSFARHDSEVDII 18  
Db 1 VPSYSFIRHDSEVDLI 18  
||||| |||||||

RESULT 9  
US-09-562-328-44  
; Sequence 44, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01

; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 44  
; LENGTH: 23  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-44

Query Match 76.4%; Score 84; DB 6; Length 23;  
Best Local Similarity 88.9%; Pred. No. 6.7e-07;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSFARHDSEVDII 18  
Db 5 VPSYSFIRHDSEVDLI 22  
||||| |||||||

RESULT 10  
US-09-604-957-5  
; Sequence 5, Application US/09604957  
; GENERAL INFORMATION:  
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
; APPLICANT: DIJKHUIZEN, LUBBERT  
; APPLICANT: RAHAOULI, HAKIM  
; APPLICANT: LEER, ROBERT-JAN  
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
; FILE REFERENCE: BO 43388  
; CURRENT APPLICATION NUMBER: US/09/604,957  
; CURRENT FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 00201871.1  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 523  
; TYPE: PRT  
; ORGANISM: Leuconostoc mesenteroides  
US-09-604-957-5

Query Match 73.6%; Score 81; DB 6; Length 523;  
Best Local Similarity 63.6%; Pred. No. 9e-05;  
Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPSYSFARHDSEVDIIRDII 22  
Db 146 IPNYSFVRHDSEVQTIVIAQIV 167  
:|:|:| ||||||| :|:|

RESULT 11  
US-09-562-328-26  
; Sequence 26, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 26  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.

US-09-562-328-26

Query Match 66.8%; Score 73.5; DB 6; Length 20;  
Best Local Similarity 84.2%; Pred. No. 2.6e-05;  
Matches 16; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

Qy 1 VPSYSFAR-AHDSEVQDII 18  
||||| | ||||| | |

Db 1 VPSYSFIRAHDSVQDLI 19

RESULT 12

US-09-562-328-31

; Sequence 31, Application US/09562328

; GENERAL INFORMATION:

; APPLICANT: LEES, ANDREW

; APPLICANT: TAUBMAN, MARTIN A.

; APPLICANT: SMITH, DANIEL J.

; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES

; FILE REFERENCE: 04995.0046-01

; CURRENT APPLICATION NUMBER: US/09/562,328

; CURRENT FILING DATE: 2000-05-01

; PRIOR APPLICATION NUMBER: 09/288,965

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 44

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 31

; LENGTH: 22

; TYPE: PRT

; ORGANISM: Streptococcus sp.

US-09-562-328-31

Query Match

62.7%; Score 69; DB 6; Length 22;

Best Local Similarity 68.2%; Pred. No. 0.00015;

Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 VPSYSFARHDSVQDIIRDII 22

||||| | ||||| | |

Db 1 VPNYVFIRAHDSVQTRIAKII 22

RESULT 13

US-09-290-049a-18

; Sequence 18, Application US/09290049A

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J.

; APPLICANT: Taubman, Martin A.

; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL

; TITLE OF INVENTION: CARIES

; FILE REFERENCE: 1564.1008-002

; CURRENT APPLICATION NUMBER: US/09/290,049A

; CURRENT FILING DATE: 1999-04-12

; PRIOR APPLICATION NUMBER: 60/081,550

; PRIOR FILING DATE: 1998-04-13

; PRIOR APPLICATION NUMBER: 60/115,142

; PRIOR FILING DATE: 1999-01-08

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 18

; LENGTH: 22

; TYPE: PRT

; ORGANISM: S. downei

US-09-290-049a-18

Query Match

62.7%; Score 69; DB 6; Length 22;

Best Local Similarity 68.2%; Pred. No. 0.00015;

Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 VPSYSFARHDSVQDIIRDII 22

||||| | ||||| | |

Db 1 VPNYVFIRAHDSVQTRIAKII 22

RESULT 14

US-09-604-957-7

; Sequence 7, Application US/09604957

; GENERAL INFORMATION:

; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA

; APPLICANT: DIJKHUIZEN, LUBBERT

; APPLICANT: RAHAOUI, HAKIM

; APPLICANT: LEER, ROBERT-JAN

; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN

; FILE REFERENCE: BO 43388

; CURRENT APPLICATION NUMBER: US/09/604,957

; CURRENT FILING DATE: 2000-06-28

; PRIOR APPLICATION NUMBER: 00201871.1

; PRIOR FILING DATE: 2000-05-25

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 7

; LENGTH: 535

; TYPE: PRT

; ORGANISM: Lactobacillus reuteri

US-09-604-957-7

Query Match

61.8%; Score 68; DB 6; Length 535;

Best Local Similarity 48.0%; Pred. No. 0.011;

Matches 12; Conservative 7; Mismatches 2; Indels 4; Gaps 1;

Qy 1 VPSYSFARAHDS----EVQDIIRDII 21

:||||| ||||: :||: |||:

Db 144 IPNYSFVRAHDNNSQDIQNAIRDV 168

RESULT 15

US-09-604-957-3

; Sequence 3, Application US/09604957

; GENERAL INFORMATION:

; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA

; APPLICANT: DIJKHUIZEN, LUBBERT

; APPLICANT: RAHAOUI, HAKIM

; APPLICANT: LEER, ROBERT-JAN

; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN

; FILE REFERENCE: BO 43388

; CURRENT APPLICATION NUMBER: US/09/604,957

; CURRENT FILING DATE: 2000-06-28

; PRIOR APPLICATION NUMBER: 00201871.1

; PRIOR FILING DATE: 2000-05-25

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3

; LENGTH: 1278

; TYPE: PRT

; ORGANISM: Lactobacillus reuteri

US-09-604-957-3

Query Match

61.8%; Score 68; DB 6; Length 1278;

Best Local Similarity 48.0%; Pred. No. 0.03;

Matches 12; Conservative 7; Mismatches 2; Indels 4; Gaps 1;

Qy 1 VPSYSFARAHDS----EVQDIIRDII 21

:||||| ||||: :||: |||:

Db 620 IPNYSFVRAHDNNSQDIQNAIRDV 644

Search completed: March 27, 2002, 14:22:50

Job time: 1697 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:01:24 ; Search time 102.51 Seconds  
(without alignments)  
16.348 Million cell updates/sec

Title: US-09-290-049a-19

Perfect score: 110

Sequence: 1 VPSYSFAHSDSEVDIIRDI 22

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Maximum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_68:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	100.0	1592	2 A38175	glucosyltransferas
2	98	89.1	1375	2 JT0345	dextranucrase (EC
3	97	88.2	1475	2 B33135	gtfB protein precu
4	85	77.3	1290	2 JC5473	dextranucrase (EC
5	82	74.5	1508	2 T31098	probable dextranu
6	72	65.5	1599	2 S22737	glucosyltransferas
7	70	63.6	1449	2 T30857	glucosyltransferas
8	70	63.6	1449	2 T30857	glucosyltransferas
9	69	62.7	1365	2 A41483	glucosyltransferas
10	67	60.9	1518	2 A44811	glucosyltransferas
11	64	58.2	1431	2 A45866	glucosyltransferas
12	64	58.2	1577	2 T30858	dextranucrase (EC
13	51	46.4	108	2 S75540	glucosyltransferas
14	49	44.5	508	2 E83788	hypothetical prote
15	46	41.8	506	2 T47184	galactose-1-phosph
16	46	41.8	540	1 OYHUCR	hypothetical prote
17	46	41.8	6359	2 T31679	baicitracin peptid
18	45	40.9	51	2 G82455	hypothetical prote
19	45	40.9	597	2 F82935	DNA polymerase III
20	44	40.0	188	2 S31626	cell fusion protei
21	44	40.0	330	2 F69471	arazine chlorohyd
22	44	40.0	1058	1 GNFF17	retrovirus-related
23	43	39.1	176	2 F83819	hypothetical prote
24	43	39.1	295	1 A64236	hypothetical prote
25	43	39.1	419	2 A40728	microphthalmia-ass
26	43	39.1	548	1 UFECAQ	fumarate hydratase
27	43	39.1	548	2 C85767	hypothetical prote
28	43	39.1	571	2 T08930	hypothetical prote
29	42.5	38.6	395	2 B75047	probable 2-oxoisov

#### ALIGNMENTS

RESULT 1

A38175

glucosyltransferase precursor - Streptococcus sobrinus

C:Species: Streptococcus sobrinus

C:Date: 28-Aug-1992 #sequence\_revision 28-Aug-1992 #text\_change 15-Oct-1999

C:Accession: A38175

R:Abou, H.; Msumura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.

J. Bacteriol. 173, 989-996, 1991

A:Title: Peptide sequences for sucrose splitting and glucan binding within Streptococ

A:Reference number: A38175; MUID:91123227

A:Accession: A38175

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1592 <ABO>

A:Cross-references: GB:D90213; NID:g217032; PIDN:BAAL4241.1; PID:d1014946; PID:g21703

C:Superfamily: cpl repeat homology

F:1093-1112/Domain: cpl repeat homology <CP1>

F:1222-1241/Domain: cpl repeat homology <CP2>

F:1287-1306/Domain: cpl repeat homology <CP3>

F:1330-1351/Domain: cpl repeat homology <CP4>

F:1352-1371/Domain: cpl repeat homology <CP5>

F:1402-1420/Domain: cpl repeat homology <CP6>

F:1465-1484/Domain: cpl repeat homology <CP7>

F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match 100.0%; Score 110; DB 2; Length 1592;

Best Local Similarity 100.0%; Pred. No. 5.4e-09;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPSYSFAHSDSEVDIIRDI 22

|||||

Db 548 VPSYSFAHSDSEVDIIRDI 569

RESULT 2

JT0345

dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)

N:Alternate names: sucrose 6-glucosyltransferase

C:Species: Streptococcus mutans

C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 24-Sep-1999

C:Accession: JT0345; C33135

R:Ueda, S.; Shiroza, T.; Kuramitsu, H.K.

Gene 69, 101-109, 1988

A:Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.

A:Reference number: JT0345; MUID:89137980

A:Accession: JT0345

A:Molecule type: DNA

A:Residues: 1-1375 <UED>

A:Experimental source: GS-5

R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

J. Bacteriol. 169, 4263-4270, 1987

A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.

A:Reference number: A33135; MUID:87308013

A:Accession: C33135

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-349 <SHI>

A:Cross-references: GB:M17361

C:Genetics:

A:Gene: gtfC

C:Function:

A:Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans

C:Superfamily: cpl repeat homology

C:Keywords: duplication; glycosyltransferase; hexosyltransferase

F:1-34/Domain: signal sequence #status predicted <SIG>

F:35-1375/Product: glycosyltransferase #status predicted <WAT>

F:1126-1145/Domain: cpl repeat homology <CP1>

F:1253-1272/Domain: cpl repeat homology <CP2>

F:1318-1337/Domain: cpl repeat homology <CP3>

Query Match 89.18; Score 98; DB 2; Length 1375;  
Best Local Similarity 86.4%; Pred. No. 3.7e-07;  
Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHDSVQDIIRDII 22  
||||| ||||| ||||| ||||| |||||  
Db 578 VPSYSFIRAHDSVQDLIRNII 599

RESULT 3

gtfB protein precursor - Streptococcus mutans

C:Species: Streptococcus mutans

C:Date: 23-Oct-1990 #sequence\_revision 23-Oct-1990 #text\_change 15-Oct-1999

C:Accession: B33135; A33128

R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

J. Bacteriol. 169, 4263-4270, 1987

A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.

A:Reference number: A33135; MUID:87308013

A:Accession: B33135

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1475 <SHI>

A:Cross-references: GB:M17361; NID:g153639; PIDN:AAA8588.1; PID:g153640

R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

submitted to the Protein Sequence Database, September 1990

Reference number: A33128

C:Accession: A33128

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-171,173-641,'N',643-1475 <SH2>

A:Experimental source: strain GS-5

C:Superfamily: cpl repeat homology

F:1096-1115/Domain: cpl repeat homology <CP1>

F:1224-1243/Domain: cpl repeat homology <CP2>

F:1289-1308/Domain: cpl repeat homology <CP3>

F:1354-1373/Domain: cpl repeat homology <CP4>

F:1419-1438/Domain: cpl repeat homology <CP5>

Query Match 88.2%; Score 97; DB 2; Length 1475;  
Best Local Similarity 86.4%; Pred. No. 5.8e-07;  
Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHDSVQDIIRDII 22  
||||| ||||| ||||| ||||| |||||  
Db 552 VPSYSFIRAHDSVQDLIRNII 573

RESULT 4

JC5473

dextranucrase (EC 2.4.1.5) - Leuconostoc mesenteroides

C:Species: Leuconostoc mesenteroides.

C:Date: 07-Jul-1997 #sequence\_revision 29-Aug-1997 #text\_change 29-Aug-1997

C:Accession: JC5473

R:Monchois, V.; Willemot, R.M.; Renaud-Simeon, M.; Croux, C.; Monsan, P.

Gene 182, 23-32, 1996

A:Title: Cloning and sequencing of a gene coding for a novel dextranucrase from Leuc

A:Reference number: JC5473; MUID:9713686

A:Accession: JC5473

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1290 <MON>

A:Cross-references: GB:U38181

C:Comment: This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose

C:Genetics:

A:Gene: dsrA

C:Keywords: glycosyltransferase; hexosyltransferase

F:78-870/Domain: catalytic #status predicted <CAT>

F:922-1290/Domain: glucan-binding #status predicted <GCB>

Query Match 77.3%; Score 85; DB 2; Length 1290;  
Best Local Similarity 81.0%; Pred. No. 4e-05;  
Matches 17; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 PSYSFARAHDSVQDIIRDII 22  
||||| ||||| ||||| ||||| |||||  
Db 388 PNYSFIRAHDSVQTIADII 408

RESULT 5

T31098

probable dextranucrase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides

C:Species: Leuconostoc mesenteroides

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000

C:Accession: T31098

R:Monchois, V.; Renaud-Simeon, M.; Monsan, P.; Willemot, R.M.

FEMS Microbiol. Lett. 159, 307-315, 1998

A:Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase

A:Reference number: Z20981; MUID:98164374

A:Accession: T31098

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-1508 <MON>

A:Cross-references: EMBL:AF030129; NID:g2766611; PID:g2766612; PIDN:AAB95453.1

A:Experimental source: strain NRRL B-1299

C:Genetics:

A:Gene: dsrB

C:Function:

A:Description: produces dextran composed only of alpha(1-6) glucosidic bonds

C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 74.5%; Score 82; DB 2; Length 1508;  
Best Local Similarity 68.2%; Pred. No. 0.00014;  
Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHDSVQDIIRDII 22  
:||| ||||| ||||| ||||| |||||  
Db 634 IPNYSFVRAHDSVQTVIAQII 655

RESULT 6

S22737

glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius

C:Species: Streptococcus salivarius

C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 21-Jan-2000

C:Accession: S22737; S28810; B4811; S22727

R:Jacques, N.

submitted to the EMBL Data Library, March 1992

A:Reference number: S22726

A:Accession: S22737

A:Molecule type: DNA

A:Residues: 1-1599 <JAC>

A:Cross-references: EMBL:Z11872; NID:g47530; PIDN:CAA77898.1; PID:g47531





C:Accession: A45866  
R:Honda, O.; Kato, C.; Kuramitsu, H.K.  
J.. Gen. Microbiol. 136, 2099-2105, 1990  
A:Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyltransferase  
A:Reference number: A45866; MUID:91100958  
A:Accession: A45866  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1431 <HON>  
C:Cross-references: GB:M29296  
C:Keywords: glycosyltransferase; hexosyltransferase  
F:161-201/Domain: cpl repeat homology <CP1>  
F:1127-1146/Domain: cpl repeat homology <CP2>  
F:1192-1211/Domain: cpl repeat homology <CP3>  
F:1257-1276/Domain: cpl repeat homology <CP4>  
F:1277-1297/Domain: cpl repeat homology <CP5>  
F:1321-1340/Domain: cpl repeat homology <CP8>  
F:1341-1361/Domain: cpl repeat homology <CP6>  
F:1385-1404/Domain: cpl repeat homology <CP7>

Query Match 58.2% Score 64; DB 2; Length 1431;  
Best Local Similarity 65.0%; Pred.No. 0.096;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 SYSFARAHSEVQDIIRDII 22  
:| | ||||| :| |  
Db 576 NYIFRAHSEVQTIAKII 595

RESULT 12  
T30858  
glucosyltransferase - Streptococcus salivarius  
C:Species: Streptococcus salivarius  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30858  
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.  
Infect. Immun. 63, 609-621, 1995  
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for proteins involved in genetic competence  
A:Reference number: Z20909; MUID:95122197  
A:Accession: T30858  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-1577 <SIM>  
C:Cross-references: EMBL:L35928; NID:g662380; PID:g662381; PIDN:AAC41413.1  
Genetics:  
Gene: gtfm

Query Match 58.2% Score 64; DB 2; Length 1577;  
Best Local Similarity 60.0%; Pred.No. 0.11;  
Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 SYSFARAHSEVQDIIRDII 22  
:| | ||||| :| |  
Db 661 NYIFRAHSEVQAVLANII 680

RESULT 13  
S75540  
hypothetical protein sl11219 - Synecchocystis sp. (strain PCC 6803)  
C:Species: Synecchocystis sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecchocystis  
A:Reference number: S74322; MUID:97061201  
A:Accession: S75540

Search completed: March 27, 2002, 14:01:25  
Job time: 487 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:26:10 ; Search time 188.53 Seconds  
(without alignments)  
17.069 Million cell updates/sec

Title: US-09-290-049a-19  
Perfect score: 110  
Sequence: 1 VPSYSFARAHDSVQDIIRDII 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL\_17:\*\*  
1: sp\_archaea:\*\*  
2: sp\_bacteria:\*\*  
3: sp\_fungi:\*\*  
4: sp\_human:\*\*  
5: sp\_invertebrate:\*\*  
6: sp\_mammal:\*\*  
7: sp\_mhc:\*\*  
8: sp\_organelle:\*\*  
9: sp\_phage:\*\*  
10: sp\_plant:\*\*  
11: sp\_rodent:\*\*  
12: sp\_virus:\*\*  
13: sp\_vertebrate:\*\*  
14: sp\_unclassified:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	110	100.0	1590	2 Q55263	Q55263 streptococc
2	110	100.0	1590	2 Q59983	Q59983 streptococc
3	103	93.6	1390	2 Q69385	Q69385 streptococc
4	103	93.6	1455	2 Q69382	Q69382 streptococc
5	103	93.6	1455	2 Q69391	Q69391 streptococc
6	103	93.6	1455	2 Q69397	Q69397 streptococc
7	98	89.1	1455	2 Q69388	Q69388 streptococc
8	85	77.3	1290	2 Q48756	Q48756 leuconostoc
9	82	74.5	1477	2 Q91466	Q91466 leuconostoc
10	82	74.5	1508	2 Q52224	Q52224 leuconostoc
11	82	74.5	1508	2 Q9EZH5	Q9EZH5 leuconostoc
12	81	73.6	1016	2 Q9LCJ7	Q9LCJ7 leuconostoc
13	81	73.6	1527	2 Q9ZAR4	Q9ZAR4 leuconostoc
14	76	69.1	1512	2 Q9WXJ5	Q9WXJ5 streptococc
15	72	65.5	1575	2 Q9LCH3	Q9LCH3 streptococc
16	72	65.5	1577	2 Q54178	Q54178 streptococc
17	72	65.5	1599	2 Q00599	Q00599 streptococc
18	70	63.6	1449	2 Q68542	Q68542 streptococc
19	70	63.6	1449	2 Q55264	Q55264 streptococc

20	69	62.7	1338	2	Q9WXJ4	Q9wxj4 streptococc
21	68	61.8	2057	2	Q9RE05	Q9re05 leuconostoc
22	67	60.9	1518	2	Q00600	Q00600 streptococc
23	64	58.2	1577	2	Q55265	Q55265 streptococc
24	51	46.4	108	2	P74028	P74028 synecocyst
25	50	45.5	93	2	Q9ZIX9	Q9zix9 borrelia bu
26	46	41.8	336	4	Q9Y3S1	Q9y3s1 homo sapien
27	46	41.8	367	4	Q9H769	Q9h769 homo sapien
28	46	41.8	506	4	Q9NSL5	Q9nsl5 homo sapien
29	46	41.8	743	4	Q9NVJ7	Q9nvj7 homo sapien
30	46	41.8	743	4	Q9NV74	Q9nv74 homo sapien
31	46	41.8	743	4	Q9BUN0	Q9bun0 homo sapien
32	46	41.8	779	4	Q9H3P4	Q9h3p4 homo sapien
33	45.5	41.4	484	2	Q99W75	Q99w75 staphylococ
34	45	40.9	51	2	Q9RNP7	Q9rnp7 vibrio chol
35	45	40.9	361	5	Q9GWP3	Q9gwp3 leishmania
36	45	40.9	597	2	Q9PR58	Q9pr58 ureaplasma
37	44	40.0	179	12	Q9Q8W3	Q9q8w3 shope fibro
38	44	40.0	188	12	Q83655	Q83655 myxoma viru
39	44	40.0	188	12	Q9Q8I5	Q9q8i5 myxoma viru
40	44	40.0	518	10	Q9FTE2	Q9fte2 oryza sativ
41	43	39.1	176	2	Q9KD61	Q9kd61 bacillus ha
42	43	39.1	377	11	O70241	O70241 mesocricetu
43	43	39.1	571	10	Q9SUA0	Q9suao arabidopsis
44	43	39.1	2627	4	Q99973	Q99973 homo sapien
45	42.5	38.6	104	2	Q9X771	Q9x771 listeria mo

ALIGNMENTS

RESULT 1  
Q55263 PRELIMINARY; PRT; 1590 AA.  
ID Q55263;  
AC Q55263;  
DT 01-JAN-1996 (TREMBLrel. 01, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE GTF-I.  
OS Streptococcus sobrinus.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID 1310;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 33478;  
RA Sato S.;  
RL Ann. Kagoshima Univ. Dental School 16:23-29(1996)  
DR EMBL; D63570; BAA09792.1;  
DR InterPro: IPR012479; CW binding.  
DR InterPro: IPR00318; Glyco\_hydro\_70.  
DR Pfam: PF01473; CW\_binding\_1; 15.  
DR Pfam: PF02327; Glyco\_hydro\_70; 1.  
KW Transferase  
SQ SEQUENCE 100 AA; 176057 MW; 9DF7A3F2G6E4FF/3 CRC64;  
  
Query Match 100.0%; Score 110; DB 2; Length 1590;  
Best Local Similarity 100.0%; Pred. No. 4.3e-08;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VPSYSFARAHDSVQDIIRDII 22  
|||||  
DB 548 VPSYSFARAHDSVQDIIRDII 569  
  
RESULT 2  
Q59983 PRELIMINARY; PRT; 1590 AA.  
ID Q59983;  
AC Q59983;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE  
DE 6-GLUCOSYLTRANSFERASE).

GN GTFC.

OS Streptococcus sobrinus.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1310;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=OM2176;

RX MEDLINE=94146405; PubMed=9312602;

RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;

RT "DNA sequence of the glucosyltransferase gene of serotype d

RT Streptococcus sobrinus.";

RL DNA Seq. 4:19-27(1993).

CC -1- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) = D-

CC FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).

DR EMBL; D13858; BAA02976.1; -

DR InterPro; IPR002479; CW\_binding.

DR Pfam; PF01473; Glyco\_hydro\_70.

DR Signal; transferase; Glycosyltransferase.

KW SIGNAL 1 38 POTENTIAL.

FT CHAIN 39 1590 GLUCOSYLTRANSFERASE-I.

SQ SEQUENCE 1590 AA; 175955 MW; C3C83A57CF3C2B0E CRC64;

Query Match 100.0%; Score 110; DB 2; Length 1590;

Best Local Similarity 100.0%; Pred. No. 4.3e-08;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPSYSFARAHSEVQDIIRDII 22

DB 548 VPSYSFARAHSEVQDIIRDII 569

RESULT 3

ID O69385 PRELIMINARY; PRT; 1390 AA.

AC O69385;

DT 01-AUG-1998 (TrEMBLrel. 07, Created)

DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE GLUCOSYLTRANSFERASE-SI.

GN GTFC.

OS Streptococcus mutans.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1309;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MT4245;

RX MEDLINE=98231643; PubMed=9570124;

RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,

RA Kimura S., Hamada S.;

RT "Molecular analyses of glucosyltransferase genes among strains of

RT Streptococcus mutans.";

RL FEMS Microbiol. Lett. 161:331-336(1998).

DR EMBL; D88655; BAA26106.1; -

DR InterPro; IPR002479; CW\_binding.

DR Pfam; PF01473; Glyco\_hydro\_70.

DR Signal; transferase; Glycosyltransferase.

KW SIGNAL 39 155375 MW; 8847E4956EF05E9F CRC64;

SQ SEQUENCE 1390 AA; 155375 MW; 8847E4956EF05E9F CRC64;

Query Match 93.6%; Score 103; DB 2; Length 1390;

Best Local Similarity 90.9%; Pred. No. 4.3e-07;

Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSFARAHSEVQDIIRDII 22

DB 548 VPSYSFARAHSEVQDIIRDII 569

Db 578 VPSYSFARAHSEVQDIIRDII 599

RESULT 4

O69382

ID O69382 PRELIMINARY; PRT; 1455 AA.

AC O69382;

DT 01-AUG-1998 (TrEMBLrel. 07, Created)

DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE GLUCOSYLTRANSFERASE-SI.

GN GTFC.

OS Streptococcus mutans.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1309;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MT8148;

RX MEDLINE=98231643; PubMed=9570124;

RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,

RA Kimura S., Hamada S.;

RT "Molecular analyses of glucosyltransferase genes among strains of

RT Streptococcus mutans.";

RL FEMS Microbiol. Lett. 161:331-336(1998).

DR EMBL; D88652; BAA26102.1; -

DR InterPro; IPR002479; CW\_binding.

DR Pfam; PF01473; Glyco\_hydro\_70.

DR Signal; transferase; Glycosyltransferase.

KW SIGNAL 39 162969 MW; 27D4D3A1EECA2939 CRC64;

SQ SEQUENCE 1455 AA; 162969 MW; 27D4D3A1EECA2939 CRC64;

Query Match 93.6%; Score 103; DB 2; Length 1455;

Best Local Similarity 90.9%; Pred. No. 4.5e-07;

Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSFARAHSEVQDIIRDII 22

DB 578 VPSYSFARAHSEVQDIIRDII 599

RESULT 5

O69391

ID O69391 PRELIMINARY; PRT; 1455 AA.

AC O69391;

DT 01-AUG-1998 (TrEMBLrel. 07, Created)

DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE GLUCOSYLTRANSFERASE-SI.

GN GTFC.

OS Streptococcus mutans.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1309;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MT4251;

RX MEDLINE=98231643; PubMed=9570124;

RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,

RA Kimura S., Hamada S.;

RT "Molecular analyses of glucosyltransferase genes among strains of

RT Streptococcus mutans.";

RL FEMS Microbiol. Lett. 161:331-336(1998).

DR EMBL; D88661; BAA26114.1; -

DR InterPro; IPR002479; CW\_binding.

DR Pfam; PF01473; Glyco\_hydro\_70.

DR Signal; transferase; Glycosyltransferase.

KW SIGNAL 39 162804 MW; 683A359D873E9E1A CRC64;

SQ SEQUENCE 1455 AA; 162804 MW; 683A359D873E9E1A CRC64;

Query Match 93.6%; Score 103; DB 2; Length 1455;  
 Best Local Similarity 90.9%; Pred. No. 4.5e-07;  
 Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VPSYSFARHDSVQDIIRDII 22  
 ||||| ||||| ||||| ||||| |||||  
 Db 578 VPSYSFIRAHDSVQDLIRNII 599

## RESULT 6

ID O69397 PRELIMINARY; PRT; 1455 AA.  
 AC O69397;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-SI.  
 GN GTFC.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MT4467;  
 RX MEDLINE=98231643; PubMed=9570124;  
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
 RA Kimura S., Hamada S.;  
 RA "Molecular analyses of glucosyltransferase genes among strains of  
 RT Streptococcus mutans";  
 RL FEMS Microbiol. Lett. 161:331-336(1998).  
 DR EMBL; D89978; BAA26120.1; -;  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding.1; 9.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase.  
 SQ SEQUENCE 1455 AA; 162913 MW; A1263427BF24E8E1 CRC64;

Query Match 93.6%; Score 103; DB 2; Length 1455;  
 Best Local Similarity 90.9%; Pred. No. 4.5e-07;  
 Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VPSYSFARHDSVQDIIRDII 22  
 ||||| ||||| ||||| ||||| |||||  
 Db 578 VPSYSFIRAHDSVQDLIRNII 599

## RESULT 7

ID O69388 PRELIMINARY; PRT; 1455 AA.  
 AC O69388;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-SI.  
 GN GTFC.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MT4239;  
 RX MEDLINE=98231643; PubMed=9570124;  
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
 RA Kimura S., Hamada S.;  
 RA "Molecular analyses of glucosyltransferase genes among strains of  
 RT Streptococcus mutans";  
 RL FEMS Microbiol. Lett. 161:331-336(1998).

DR EMBL; D88658; BAA26110.1; -;  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding.1; 10.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase.  
 SQ SEQUENCE 1455 AA; 163045 MW; 6D90A4978D35DD82 CRC64;

Query Match 89.1%; Score 98; DB 2; Length 1455;  
 Best Local Similarity 86.4%; Pred. No. 2.6e-06;  
 Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VPSYSFARHDSVQDIIRDII 22  
 ||||| ||||| ||||| ||||| |||||  
 Db 578 VPSYSFIRAHDSVQDLIRNII 599

## RESULT 8

ID Q48756 PRELIMINARY; PRT; 1290 AA.  
 AC Q48756;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE DEXTRANSUCRASE.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL B1299;  
 RX MEDLINE=97136686; PubMed=8982063;  
 RA Monchois V., Willemot R.M., Remaud-Simeon M., Croux C., Monsan P.;  
 RT "Cloning and sequencing of a gene coding for a novel dextransucrase  
 from Leuconostoc mesenteroides NRRL B-1299 synthesizing only alpha (1-  
 RT 6) and alpha (1-3) linkages";  
 RL Gene 182:23-32(1996).  
 DR EMBL; U38181; ABA40875.1; -;  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding.1; 11.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 SQ SEQUENCE 1290 AA; 145590 MW; 3555C2E96B749FAA CRC64;

Query Match 77.3%; Score 85; DB 2; Length 1290;  
 Best Local Similarity 81.0%; Pred. No. 0.00022;  
 Matches 17; Conservative 1; Mismatches 3; Indels 0; Gaps 0

Qy 2 PSYSFARHDSVQDIIRDII 22  
 ||||| ||||| ||||| ||||| |||||  
 Db 388 PSYSFIRAHDSVQDLIRNII 408

## RESULT 9

ID Q9L466 PRELIMINARY; PRT; 1477 AA.  
 AC Q9L466;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE DEXTRANSUCRASE (EC 2.4.1.5).  
 GN DSRC.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL B-1355;  
 RA Arguello-Morales M.A., Remaud-Simeon M., Pizzut S., Sarcabal P.,

RA Willemot R.M., Monsan P.;  
 RT "Sequence analysis of the gene encoding alternansucrase, a sucrose  
 RL glucosyltransferase from *Leuconostoc mesenteroides* NRRL B-1355.";  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ250172; CAB65651.1; -;  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 14.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Transferrase; Glycosyltransferase.  
 SQ SEQUENCE 1477 AA; 16486 MW; E6F5710DEDFCB831 CRC64;

Query Match 74.5%; Score 82; DB 2; Length 1477;  
 Best Local Similarity 68.2%; Pred. No. 0.00074;  
 Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Oy 1 VPSYFARAHDSVQDIIRDII 22  
 Db 603 IPNYSFVRAHDSVQTVIAQII 624  
 :||||| ||||| :|||

RESULT 10  
 O52224  
 ID O52224 PRELIMINARY; PRT; 1508 AA.  
 AC O52224;  
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE 6-  
 DE GLUCOSYLTRANSFERASE).  
 GN DSRB.  
 OS *Leuconostoc mesenteroides*.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL B-1299;  
 RA Monchois V., Renaud-Simeon M., Monsan P., Willemot R.M.;  
 RL FEMS Microbiol. Lett. 0:0-0(1998).  
 CC -1- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) = D-  
 CC FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).  
 DR EMBL; AF030129; AAB95453.1; -;  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 14.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Transferrase; Glycosyltransferase.  
 SQ SEQUENCE 1508 AA; 168511 MW; E70CECB57A70D1F0 CRC64;

Query Match 74.5%; Score 82; DB 2; Length 1508;  
 Best Local Similarity 68.2%; Pred. No. 0.00076;  
 Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Oy 1 VPSYFARAHDSVQDIIRDII 22  
 Db 634 IPNYSFVRAHDSVQTVIAQII 655  
 :||||| ||||| :|||

RESULT\* 11  
 O9EZH5  
 ID O9EZH5 PRELIMINARY; PRT; 1508 AA.  
 AC O9EZH5;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE DEXTRANSUCRASE DSRB742.  
 GN DSRB742.  
 OS *Leuconostoc mesenteroides*.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC Leuconostoc.

OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B-742CB;  
 RA Kim H.-S., Kim D., Ryu H.-J., Robyt J.F.;  
 RT "Leuconostoc mesenteroides B-742CB, a dextranucrase gene.";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF294469; AAG38021.1; -;  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 14.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 SQ SEQUENCE 1508 AA; 168542 MW; E2FCFA0F87AE4F3A CRC64;

Query Match 74.5%; Score 82; DB 2; Length 1508;  
 Best Local Similarity 68.2%; Pred. No. 0.00076;  
 Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Oy 1 VPSYFARAHDSVQDIIRDII 22  
 Db 634 IPNYSFVRAHDSVQTVIAQII 655  
 :||||| ||||| :|||

RESULT 12  
 O9LCJ7  
 ID O9LCJ7 PRELIMINARY; PRT; 1016 AA.  
 AC O9LCJ7;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE DEXTRANSUCRASE.  
 GN DSRT.  
 OS *Leuconostoc mesenteroides*.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL B-512F;  
 RA MEDLINE=20169623; PubMed=10705445;  
 RA Funane K., Mizuno K., Takahara H., Kobayashi M.;  
 RT "Gene encoding a dextranucrase-like protein in *Leuconostoc*  
 RT *mesenteroides* NRRL B-512F.";  
 RL Biosci. Biotechnol. Biochem. 64:29-38(2000).  
 DR EMBL; AB020020; BAA90527.1; -;  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 SQ SEQUENCE 1016 AA; 110343 MW; 8896EFD13CCCB47 CRC64;

Query Match 73.6%; Score 81; DB 2; Length 1016;  
 Best Local Similarity 71.4%; Pred. No. 0.0007;  
 Matches 15; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Oy 2 PSYFARAHDSVQDIIRDII 22  
 Db 625 PNYSFVRAHDSVQTVIAEII 645  
 :||||| ||||| :|||

RESULT 13  
 O9ZAR4  
 ID O9ZAR4 PRELIMINARY; PRT; 1527 AA.  
 AC O9ZAR4;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE DEXTRANSUCRASE.  
 GN DEX.  
 OS *Leuconostoc mesenteroides*.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC Leuconostoc.  
 OX NCBI\_TaxID=1245;



```

RN  SEQUENCE FROM N.A.
RP  STRAIN=NRRL B-512-F;
RC  "Cloning and Molecular Characterization of Dextranucrase Gene from
RT  Bhatnagar R., Singh D.K.S.;
RL  Leuconostoc mesenteroides NRRL B-512F.";
RM  Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR  EMBL; U81374; AAD10952.1;
DR  InterPro; IPR002479; CW_binding.
DR  InterPro; IPR003318; Glyco_hydro_70.
DR  Pfam; PF01473; CW_binding_1; 16.
DR  Pfam; PF02324; Glyco_hydro_70; 1.
SQ  SEQUENCE 1527 AA; 169709 MW; 1DFAFA237C743398 CRC64;

Query Match 73.6%; Score 81; DB 2; Length 1527;
Best Local Similarity 63.6%; Pred. No. 0.0011;
Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 VPSYSFAHDSFVQDIIRDII 22
    :||||| ||||| :| :|
Db 652 IPNTSFVRAHDSFVQTIVIAQIV 673

RESULT 14
ID  Q9WXJ5 PRELIMINARY; PRT; 1512 AA.
AC  Q9WXJ5;
DT  01-NOV-1999 (TrEMBLrel. 12, Created)
DT  01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT  01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE  GTF-S.
GN  GTF.
OS  Streptococcus criceti.
OC  Plasmid pAM1
OC  Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC  Streptococcus.
OX  NCBI_TaxID=1333;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=HS-6;
RA  Inoue M., Fukui K., Miyagi A.;
RT  "S.cricetus glucosyltransferase(gtfs and gtfT) genes.";
RL  Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AB026123; BAA77237.1;
DR  InterPro; IPR02479; CW_binding.
DR  InterPro; IPR003318; Glyco_hydro_70.
DR  Pfam; PF01473; CW_binding_1; 14.
DR  Pfam; PF02324; Glyco_hydro_70; 1.
KW  Plasmid.
SQ  SEQUENCE 1512 AA; 167145 MW; 4C03D9C6C601FC14 CRC64;

Query Match 69.1%; Score 76; DB 2; Length 1512;
Best Local Similarity 71.4%; Pred. No. 0.0063;
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 2 PSYSFAHDSFVQDIIRDII 22
    ||| | ||||| :| ||
Db 560 PSYFVRAHDSFVQTIVIAQII 580

RESULT 15
Q9LCH3 PRELIMINARY; PRT; 1575 AA.
ID  Q9LCH3
AC  Q9LCH3;
DT  01-OCT-2000 (TrEMBLrel. 15, Created)
DT  01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT  01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE  GLUCOSYLTRANSFERASE.
GN  GTF.
OS  Streptococcus oralis.
OC  Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

```

```

OC  Streptococcus.
OX  NCBI_TaxID=1303;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=ATCC10557;
RX  MEDLINE=20231779; PubMed=10768934;
RA  Fujiwara T., Hoshino T., Ooshima T., Sobue S., Hanada S.;
RT  "Purification, characterization, and molecular analysis of the gene
RL  encoding glucosyltransferase from Streptococcus oralis.";
DR  Infect. Immun. 68:2475-2483(2000).
DR  EMBL; AB025228; BAA95201.1;
DR  InterPro; IPR002479; CW_binding.
DR  InterPro; IPR003318; Glyco_hydro_70.
DR  Pfam; PF01473; CW_binding_1; 17.
DR  Pfam; PF02324; Glyco_hydro_70; 1.
KW  Transferase.
SQ  SEQUENCE 1575 AA; 176792 MW; 772A26EAD7C2E543 CRC64;

Query Match 65.5%; Score 72; DB 2; Length 1575;
Best Local Similarity 70.0%; Pred. No. 0.027;
Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 3 SYSFAHDSFVQDIIRDII 22
    :| | ||||| :| |||
Db 619 NYFVRAHDSFVQTIVIAQII 638

Search completed: March 27, 2002, 14:26:11
Job time: 1683 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:27:08 ; Search time 53.4 seconds  
(without alignments)  
15.105 Million cell updates/sec

Title: US-09-290-049a-19  
Perfect score: 110  
Sequence: 1 VPSYFARAHSDSEVDIIRDI 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	110	100.0	1592	1 GTF2_STRDO	P27470 streptococc
2	108	98.2	1597	1 GTF1_STRDO	P11001 streptococc
3	103	93.6	1476	1 GTFB_STRMU	P08987 streptococc
4	98	89.1	1375	1 GTFB_STRMU	P13470 streptococc
5	69	62.7	1365	1 GTFB_STRMU	P29336 streptococc
6	64	58.2	1462	1 GTFD_STRMU	P49331 streptococc
7	49	44.5	508	1 GAL7_BACHD	Q9KQV2 bacillus ha
8	46	41.8	540	1 ANPC_HUMAN	P17342 homo sapien
9	46	41.8	6359	1 BACC_BACLI	O68008 b bacitraci
10	44	40.0	330	1 YH75_ARCFU	O28499 archaeoglob
11	44	40.0	1046	1 RPOC_WEIHE	P96177 weissella h
12	44	40.0	1058	1 POL3_DROME	P04323 drosophila
13	43	39.1	295	1 Y326_MYCCE	P47568 mycoplasma
14	43	39.1	526	1 MITF_MOUSE	Q08874 mus musculu
15	43	39.1	547	1 FUMA_ECOLI	P00923 escherichia
16	42.5	38.6	395	1 VORA_PYRAB	Q9UYZ1 pyrococcus
17	42.5	38.6	494	1 SYE_MYCPN	P75114 mycoplasma
18	42.5	38.6	540	1 HXTD_YEAST	P42833 saccharomyc
19	42	38.2	196	1 MAD2_YEAST	P40958 saccharomyc
20	42	38.2	537	1 ANPC_BOVIN	P10730 bos taurus
21	42	38.2	1389	1 PRAX_RAT	O63425 rattus norv
22	41	37.3	109	1 CYC6_CYACA	Q9TLW1 cyanidium c
23	41	37.3	280	1 YIBQ_HAEIN	P44863 haemophilus
24	41	37.3	306	1 FMRF_LYMYT	P19802 lymnaea sta
25	41	37.3	414	1 FTZ2_PYRHO	O58491 pyrococcus
26	41	37.3	453	1 MDHP_FLABI	P46489 flaveria bi
27	41	37.3	691	1 YKIO_YEAST	P40460 saccharomyc
28	40	36.4	169	1 YKHO_YEAST	P36087 saccharomyc
29	40	36.4	207	1 KUQA_THEMEA	Q9X215 thermotoga
30	40	36.4	211	1 UL92_HSVJ	P52471 human heipe
31	40	36.4	314	1 TOP1_SFVKA	P16472 Shope fibro
32	40	36.4	344	1 M12D_BACSU	P26935 bacillus su
33	40	36.4	429	1 TF3A_YEAST	P39933 saccharomyc

34 40 36.4 495 1 Y892\_MYCTU Q10532 mycobacteri  
35 40 36.4 659 1 RGN\_HAEIN P44440 haemophilus  
36 40 36.4 1018 1 VGNM\_BPMV P23009 bean-pod mo  
37 40 36.4 1068 1 P11A\_BOVIN P32871 bos taurus  
38 40 36.4 1068 1 P11A\_MOUSE P42336 homo sapien  
39 40 36.4 1068 1 P11A\_MOUSE P32528 mus musculu  
40 40 36.4 1835 1 DURL\_YEAST P32528 saccharomyc  
41 39.5 35.9 183 1 RETB\_BOVIN P18902 bos taurus  
42 39.5 35.9 201 1 RETB\_PIG P27485 sus scrofa  
43 39.5 35.9 297 1 ARGE\_SINY3 P73326 synechocyst  
44 39 35.5 96 1 IBBA\_PEA Q41065 pisum sativ  
45 39 35.5 114 1 IBB2\_PEA Q41066 pisum sativ

## ALIGNMENTS

RESULT 1  
GTF2\_STRDO  
ID GTF2\_STRDO STANDARD; PRT; 1592 AA.  
AC P27470;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)  
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
OS Streptococcus downei (Streptococcus sobrinus).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1317;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=6715;  
RX MEDLINE=91123227; PubMed=1704006;  
RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,  
RA Kagawa H.;  
RT "Peptide sequences for sucrose splitting and glucan binding within  
RT Streptococcus sobrinus glucosyltransferase (water-insoluble glucan  
RT synthetase)";  
RL J. Bacteriol. 173:989-996(1991).  
CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -!- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -  
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
CC -!- SUBCELLULAR LOCATION: SECRETED.  
CC -!- DISEASE: DENTAL CARIES.  
CC -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
CC 1,3-LINKED GLUCOSE AND SOME 1,6-LINKAGES), GTF-S SYNTHESIZES  
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
CC FORMS OF GLUCANS.  
CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
CC BINDING PROTEIN FROM S.MUTANS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
-----  
DR EMBL: D90213; BAA14241.1; -  
DR PIR: A38175; A38175.  
DR HSSP: P00695; 2HEE.  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam: PF01473; CW\_binding\_1; 16.  
DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
DR Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.  
FT SIGNAL 1 38 POTENTIAL.  
FT CHAIN 39 1592 GLUCOSYLTRANSFERASE-I.

FT DOMAIN 39 1044 CATALYTIC (APPROXIMATE).  
 FT DOMAIN 1093 1592 GLUCAN-BINDING (APPROXIMATE).  
 FT DOMAIN 1093 1592 6.5 X TANDEM REPEATS.  
 FT REPEAT 1093 1142 1.  
 FT REPEAT 1158 1207 2.  
 FT REPEAT 1222 1272 3.  
 FT REPEAT 1287 1337 4.  
 FT REPEAT 1402 1451 5.  
 FT REPEAT 1514 1563 6.  
 FT REPEAT 1577 1592 7 (INCOMPLETE).  
 SQ SEQUENCE 1592 AA; 176167 MW; BC0A66D079351ECF CRC64;

Query Match 100.0%; Score 110; DB 1; Length 1592;  
 Best Local Similarity 100.0%; Pred. No. 5.3e-09; Indels 0; Gaps 0;  
 Matches 22; Conservative 0; Mismatches 0;

Qy 1 VPSYSFARHDSEVDIIRDI 22  
 Db 548 VPSYSFARHDSEVDIIRDI 569

## RESULT 2

GTFL\_STRDO STANDARD; PRT; 1597 AA.  
 AC P11001;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)  
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
 GN GTFI.  
 OS Streptococcus downei (Streptococcus sobrinus).  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1317;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MFE28;  
 RA MEDLINE=87308014; PubMed=3040686;  
 RA Ferretti J.J., Gilpin M.L., Russell R.R.B.;  
 RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus sobrinus MFE28";  
 RL J. Bacteriol. 169:4271-4278(1987).  
 CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -!- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) = D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -!- SUBCELLULAR LOCATION: SECRETED.  
 CC -!- DISEASE: DENTAL CARIES.  
 CC -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF GLUCANS.  
 CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-BINDING PROTEIN FROM S. MUTANS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M17391; AAC63063.1; -  
 DR InterPro; IPR002479; CW.binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 19.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW \* Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.

FT SIGNAL 1 38 POTENTIAL.  
 FT CHAIN 39 1597 GLUCOSYLTRANSFERASE-I.  
 FT DOMAIN 39 1050 CATALYTIC (APPROXIMATE).  
 FT DOMAIN 1099 1597 GLUCAN-BINDING (APPROXIMATE).  
 FT DOMAIN 1099 1597 1.25 A, 2 B, AND 5 AC REPEATS.  
 FT REPEAT 1099 1132 A REPEAT.  
 FT REPEAT 1163 1213 AC REPEAT.  
 FT REPEAT 1227 1277 AC REPEAT.  
 FT REPEAT 1292 1342 AC REPEAT.  
 FT REPEAT 1352 1399 B REPEAT.  
 FT REPEAT 1406 1455 AC REPEAT.  
 FT REPEAT 1465 1512 B REPEAT.  
 FT REPEAT 1519 1568 AC REPEAT.  
 FT REPEAT 1582 1597 A REPEAT (INCOMPLETE).  
 SQ SEQUENCE 1597 AA; 177080 MW; B9E86A200868798E CRC64;

Query Match 98.2%; Score 108; DB 1; Length 1597;  
 Best Local Similarity 95.5%; Pred. No. 1.1e-08;  
 Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPSYSFARHDSEVDIIRDI 22  
 Db 554 VPSYSFARHDSEVDIIRDI 575

## RESULT 3

GTFB\_STRMU STANDARD; PRT; 1476 AA.  
 AC P08987; O69381; O69384; O69387; O69390; O69396;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)  
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
 GN GTFB.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GS-5;  
 RX MEDLINE=87308013; PubMed=3040685;  
 RA Shiroza T., Ueda S., Kuramitsu H.K.;  
 RT "Sequence analysis of the gtfB gene from Streptococcus mutans";  
 RL J. Bacteriol. 169:4263-4270(1987).  
 CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -!- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) = D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -!- SUBCELLULAR LOCATION: SECRETED.  
 CC -!- DISEASE: DENTAL CARIES.  
 CC -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF GLUCANS.  
 CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-BINDING PROTEIN FROM S. MUTANS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M17391; AAC63063.1; -  
 DR InterPro; IPR002479; CW.binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 19.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW \* Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

EMBL; M17361; AAA88588.1; -
EMBL; D88651; BAA26101.1; -
EMBL; D88654; BAA26105.1; -
EMBL; D88657; BAA26109.1; -
EMBL; D88660; BAA26113.1; -
EMBL; D89977; BAA26119.1; -
PIR; B33135; B33135.
InterPro; IPR002479; CW_binding.
Pfam; PF01473; Glyco_hydro_70.
Pfam; PF02324; Glyco_hydro_70; 1.
Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
SIGNA
CHAIN 1 34
DOMAIN 35 1476
DOMAIN 35 1051
REPEAT 1097 1476
REPEAT 1097 1130
REPEAT 1161 1470
REPEAT 1161 1210
REPEAT 1225 1275
REPEAT 1290 1340
REPEAT 1355 1405
REPEAT 1420 1470
VARIANT 62 62
VARIANT 65 65
VARIANT 68 68
VARIANT 78 78
VARIANT 86 86
VARIANT 89 89
VARIANT 168 168
VARIANT 276 276
VARIANT 399 399
VARIANT 474 474
VARIANT 512 512
VARIANT 519 519
VARIANT 701 701
VARIANT 708 708
VARIANT 938 938
VARIANT 952 957
VARIANT 963 964
VARIANT 968 970
VARIANT 1086 1086
VARIANT 1158 1158
VARIANT 1163 1163
VARIANT 1168 1168
VARIANT 1182 1182
VARIANT 1234 1234
VARIANT 1263 1263
VARIANT 1263 1263
VARIANT 1264 1264
VARIANT 1272 1272
VARIANT 1329 1329
VARIANT 1394 1394
VARIANT 1402 1402
VARIANT 1459 1459
CONFLICT 570 570
CONFLICT 800 817
CONFLICT 1310 1310
SEQUENCE 1476 AA; 165685 MW; 3479B62B07694D98 CRC64;

```

Query Match 93.6%; Score 103; DB 1; Length 1476;  
 Best Local Similarity 90.9%; Pred. No. 5.9e-08;  
 Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSFARAHSEVQDIIRDII 22  
 ||||| ||||| ||||| |||||  
 Db 552 VPSYSFIRAHSEVQDLIRDI 573

## RESULT 4

GTFC\_STRMU STANDARD; PRT; 1375 AA.  
 AC PI3470; P05427;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-SI PRECURSOR (EC 2.4.1.5) (GTF-SI)  
 DE (DEXTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE).  
 GN GTFC.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OC NCBI\_TaxID=1309;  
 [1]  
 SEQUENCE FROM N.A.  
 RP STRAIN=GS-5;  
 RC MEDLINE=89137980; PubMed=2976010;  
 RX Ueda S., Shiroza T., Kuramitsu H.K.;  
 RA "Sequence analysis of the gtfc gene from Streptococcus mutans GS-5.";  
 RT Gene 69:101-109(1988).  
 [2]  
 SEQUENCE OF 1-349 FROM N.A.  
 RP STRAIN=GS-5;  
 RC MEDLINE=87308013; PubMed=3040685;  
 RX Shiroza T., Ueda S., Kuramitsu H.K.;  
 RA "Sequence analysis of the gtfc gene from Streptococcus mutans.";  
 RT J. Bacteriol. 169:4263-4270(1987).  
 CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE  
 CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -!- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) =  
 CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).  
 CC -!- SUBCELLULAR LOCATION: SECRETED.  
 CC -!- DISEASE: DENTAL CARIES.  
 CC -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
 CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
 CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
 CC FORMS OF GLUCANS.  
 CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
 CC BINDING PROTEIN FROM S.MUTANS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M22054; AAA88592.1; -  
 EMBL; M17361; AAA88589.1; -  
 PIR; JTO345; JTO345.  
 PIR; C33135; C33135.  
 InterPro; IPR002479; CW\_binding.  
 InterPro; IPR003318; Glyco\_hydro\_70.  
 Pfam; PF01473; CW\_binding\_1; 7.  
 Pfam; PF02324; Glyco\_hydro\_70; 1.  
 Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.  
 FT SIGNAL 1 34  
 FT CHAIN 35 1375 GLUCOSYLTRANSFERASE-SI.



```
CC DR EMBL; M29296; AAA26895.1; -
DR EMBL; D88653; BAA26103.1; -
DR EMBL; D88656; BAA26107.1; -
DR EMBL; D88659; BAA26111.1; -
DR EMBL; D88662; BAA26115.1; -
DR EMBL; D89979; BAA26121.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding.1; 11.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1 ?
FT CHAIN ? 1462
FT DOMAIN 1232 1423
FT REPEAT 1232 1295
FT REPEAT 1296 1359
FT REPEAT 1360 1423
FT VARIANT 58 58
FT VARIANT 68 68
FT VARIANT 81 81
FT VARIANT 113 113
FT VARIANT 122 122
FT VARIANT 132 132
FT VARIANT 135 135
FT VARIANT 202 202
FT VARIANT 255 255
FT VARIANT 275 275
FT VARIANT 288 288
FT VARIANT 301 301
FT VARIANT 313 313
FT VARIANT 317 317
FT VARIANT 328 328
FT VARIANT 350 350
FT VARIANT 628 633
FT VARIANT 688 688
FT VARIANT 726 732
FT VARIANT 726 730
FT VARIANT 964 964
FT VARIANT 1019 1019
FT VARIANT 1059 1060
FT VARIANT 1060 1060
FT VARIANT 1080 1080
FT VARIANT 1142 1142
FT VARIANT 1198 1198
FT VARIANT 1220 1220
FT VARIANT 1280 1280
FT VARIANT 1282 1282
FT VARIANT 1290 1290
FT VARIANT 1311 1311
FT VARIANT 1403 1403
FT VARIANT 1425 1425
FT VARIANT 1449 1449
FT CONFLICT 1428 1462
SQ SEQUENCE 1462 AA; 163512 MW; 5C6541F0DCB0DF00 CRC64;

Query Match 58.28; Score 64; DB 1; Length 1462;
Best Local Similarity 65.08; Pred. No. 0.063;
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 3 SYSTRANDSEVQDIIRDII 22
:| | | | | | | | | | | | | |
```

```
Db 576 NYIFRAHDSVQTVIAKII 595
RESULT 7
GAL7_BACHD STANDARD; PRT; 508 AA.
ID GAL7_BACHD
AC Q9KDV2;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GALACTOSE-1-PHOSPHATE URIDYLTRANSFERASE (EC 2.7.7.10) (GAL-1-P
URIDYLTRANSFERASE).
GN GALT OR BH1109.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RA *Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -1- CATALYTIC ACTIVITY: UTP + ALPHA-D-GALACTOSE 1-PHOSPHATE =
CC UDP-GALACTOSE + PYROPHOSPHATE.
CC -1- PATHWAY: GALACTOSE METABOLISM.
CC -1- SIMILARITY: BELONGS TO THE GALACTOSE-1-PHOSPHATE
CC URIDYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP001510; BAB04828.1; -
CC InterPro; IPR000880; GalP_UDP_transf.
CC InterPro; IPR000766; GalP_UDP_transf_II.
CC Pfam; PF01087; GalP_UDP_transf; 1.
CC PROSITE; PS01163; GAL_P_UDP_TRANSF_II; 1.
CC Transferase; Nucleotidyltransferase; Galactose metabolism;
CC Complete proteome.
CC SEQUENCE 508 AA; 57989 MW; 16AF6F607FCEAE2E CRC64;

Query Match 44.58; Score 49; DB 1; Length 508;
Best Local Similarity 62.58; Pred. No. 4.1;
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 7 ARAHDSVQDIIRDII 22
| | | | | | | | | |
Db 460 AHLHDSNVMDILRDEI 475

RESULT 8
ANPC_HUMAN STANDARD; PRT; 540 AA.
ID ANPC_HUMAN
AC P17342;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ATRIAL NATRIURETIC PEPTIDE CLEARANCE RECEPTOR PRECURSOR (ANP-C)
DE (ANPRC) (NPR-C) (ATRIAL NATRIURETIC PEPTIDE C-TYPE RECEPTOR).
GN NPR3 OR ANPRC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney;  
RX MEDLINE=90287735; PubMed=2162522;  
RA Lowe D.G., Camerato T.R., Goeddel D.V.;  
RT "cDNA sequence of the human atrial natriuretic peptide clearance  
receptor";  
RL Nucleic Acids Res. 18:3412-3412(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9038656; PubMed=2169733;  
RA Porter J.G., Arfsten A., Fuller J.A., Gregory L.C.,  
RA Lewicki J.A.;  
RT "Isolation and functional expression of the human atrial natriuretic  
peptide clearance receptor cDNA";  
RL Biochem. Biophys. Res. Commun. 171:796-803(1990).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lens epithelium;  
RA Rae J.L., Shepard A.R.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: RECEPTOR FOR ATRIAL NATRIURETIC PEPTIDE. DOES NOT HAVE  
GUANYLATE CYCLASE ACTIVITY.  
CC -!- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -!- MISCELLANEOUS: THERE SEEM TO BE AT LEAST THREE ANP RECEPTORS: TWO  
WITH GUANYLATE CYCLASE ACTIVITY (ANP-A AND ANP-B) AND ONE (ANP-C)  
WHICH IS PROBABLY RESPONSIBLE FOR THE CLEARANCE OF ANP FROM THE  
CIRCULATION WITHOUT A ROLE IN SIGNAL TRANSDUCTION.  
CC -!- SIMILARITY: TO ANP-A AND ANP-B RECEPTORS IN THEIR EXTRACELLULAR  
AND TRANSMEMBRANE DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X5282; CAA36523.1; -  
DR EMBL; M59305; AAA51734.1; -  
DR EMBL; AF025998; AAB88801.1; -  
DR PIR; S10150; S10150.  
DR PIR; A35896; A35896.  
DR MIM; 108962; -  
DR InterPro; IPR001170; ANP\_receptor.  
DR InterPro; IPR001828; ANF\_receptor.  
DR Pfam; PF01094; ANF\_receptor; 1.  
DR PRINTS; PR00255; NATPEPTIDER.  
DR PROSITE; PS00458; ANP\_RECEPTORS; 1.  
DR Receptor; Glycoprotein; Transmembrane; Signal.  
FT SIGNAL 1  
FT CHAIN 18  
FT CHAIN 19 540 ATRIAL NATRIURETIC PEPTIDE CLEARANCE  
FT RECEPTOR.  
FT DOMAIN 19 480 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 481 503 POTENTIAL.  
FT DOMAIN 504 540 CYTOPLASMIC (POTENTIAL).  
FT DISULFID 108 136 BY SIMILARITY.  
FT DISULFID 213 261 BY SIMILARITY.  
FT DISULFID 473 473 INTERCHAIN (BY SIMILARITY).  
FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 476 476 C -> SG (IN REF. 2).  
SQ SEQUENCE 540 AA; 59766 MW; 53EE0020A296D6F5 CRC64;

Query Match

Best Local Similarity 41.8%; Score 46; DB 1; Length 540;

Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 4 YSFARHDSVQDIIRDI 21  
||| | : : : | : | : |  
Db 234 YSFDTRKDLLEDIVRNI 251  
RESULT 9  
BACC\_BACLI  
ID BACC\_BACLI STANDARD; PRT; 6359 AA.  
AC O68008;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE BACITRACIN SYNTHETASE 3 (BA3) [INCLUDES: ATP-DEPENDENT ISOLEUCINE  
ADENYLASE (ILEA) (ISOLEUCINE ACTIVASE); ATP-DEPENDENT D-PHENYLALANINE  
ADENYLASE (D-PHEA) (D-PHENYLALANINE ACTIVASE); ATP-DEPENDENT HISTIDINE  
ADENYLASE (HISA) (HISTIDINE ACTIVASE); ATP-DEPENDENT D-ASPARTATE  
ADENYLASE (D-ASPA) (D-ASPARTATE ACTIVASE); ATP-DEPENDENT ASPARAGINE  
ADENYLASE (ASNA) (ASPARAGINE ACTIVASE); ASPARTATE RACEMASE  
(EC 5.1.1.13); PHENYLALANINE RACEMASE [ATP HYDROLYSING]  
(DE EC 5.1.1.11)].  
DE BACC.  
GN Bacillus licheniformis.  
OS Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OC NCBI\_TaxID=1402;  
CC [1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=ATCC 10716;  
RC MEDLINE=9808193; PubMed=9427658;  
RA Konz D., Klenz A., Schoegendorfer K., Marahiel M.A.;  
RT "The bacitracin biosynthesis operon of Bacillus licheniformis ATCC  
10716: molecular characterization of three multi-modular peptide  
synthetases";  
RL Chem. Biol. 4:927-937(1997).  
CC -!- FUNCTION: INDUCES PEPTIDE SYNTHESIS, ACTIVATES AND INCORPORATES  
FIVE AMINO ACIDS, FORMS A THIAZOLINE RING BETWEEN THE FIRST TWO  
AMINO ACIDS AND INCORPORATES A D-GLUTAMINE IN THE FOURTH POSITION.  
CC -!- CATALYTIC ACTIVITY: L-ASPARTATE -> D-ASPARTATE.  
CC -!- CATALYTIC ACTIVITY: ATP + L-PHENYLALANINE -> AMP + PYROPHOSPHATE  
+ D-PHENYLALANINE.  
CC -!- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTETHEINES  
(POTENTIAL).  
CC -!- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF THE CYCLIC PEPTIDE  
ANTIBIOTIC BACITRACIN.  
CC -!- SUBUNIT: LARGE MULTIZENZYME COMPLEX OF BA1, BA2 AND BA3.  
CC -!- DOMAIN: CONSISTS OF FIVE MODULES WITH TWO EPIMERIZATION DOMAINS IN  
THE SECOND AND FOURTH MODULES, AND A PUTATIVE C-TERMINAL  
THIOESTERASE DOMAIN. EACH MODULE INCORPORATES ONE AMINO ACID INTO  
THE PEPTIDE PRODUCT AND CAN BE FURTHER SUBDIVIDED INTO DOMAINS  
RESPONSIBLE FOR SUBSTRATE ADENYLATION, THIOATION, CONDENSATION  
(NOT FOR THE INITIATION MODULE), AND EPIMERIZATION (OPTIONAL), AND  
N METHYLATION (OPTIONAL).  
CC -!- MISCELLANEOUS: BACITRACIN IS A MIXTURE OF AT LEAST TEN CYCLIC  
DODECAPEPTIDES, THAT DIFFER BY ONE OR TWO AMINO ACIDS. THE MOST  
ABUNDANT IS BACITRACIN A, A BRANCHED CYCLIC DODECAPEPTIDE. IT  
CONTAINS AN N-TERMINAL LINEAR PENTAPEPTIDE MOIETY (ILE-CYS-LEU-D-  
GLU-ILE) WITH AN ISOLEUCINE-CYSTEINE THIAZOLINE CONDENSATION  
PRODUCT AND A C-TERMINAL HEPTAPEPTIDE RING (LYS-D-ORN-ILE-D-PHE-  
HIS-D-ASP-ASN), IN WHICH THE FREE ALPHA-CARBOXY GROUP OF THE C-  
TERMINAL ASN IS BOUND TO THE EPSILON-AMINO GROUP OF LYSINE. IT  
CONTAINS FOUR AMINO ACIDS IN THE D-CONFIGURATION (GLU-4, ORN-7,  
PHE-9, AND ASP-11).  
CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME  
FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----



```

DR EMBL; AF007865; AAC06348.1; -.
DR InterPro; IPR000873; AMP-Bind.
DR InterPro; IPR001242; DUF4.
DR InterPro; IPR003880; Phosphopant_attach.
DR InterPro; IPR001031; Thioesterase.
DR Pfam; PF00501; AMP-binding; 5.
DR Pfam; PF00668; Condensation; 7.
DR Pfam; PF00550; pp-binding; 5.
DR Pfam; PF00975; Thioesterase; 1.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 4.
DR PROSITE; PS00455; AMP_BINDING; 5.
DR PROSITE; PS00075; ACP_DOMAIN; 5.
KW Ligase; Isomerase; Hydrolase; Antibiotic biosynthesis;
KW Phosphopantetheine; Multifunctional enzyme; Repeat.
FT REPEAT 461 1034 DOMAIN 1 (ISOLEUCINE-ACTIVATING).
FT REPEAT 1517 2064 DOMAIN 2 (D-PHENYLALANINE-ACTIVATING).
FT REPEAT 2399 3570 DOMAIN 3 (HISTIDINE-ACTIVATING).
FT REPEAT 4047 4612 DOMAIN 4 (D-ASPARTIC ACID-ACTIVATING).
FT REPEAT 5549 6129 DOMAIN 5 (ASPARAGINE-ACTIVATING).
FT DOMAIN 966 1034 ACYL CARRIER (ACP) 1.
FT DOMAIN 1998 2064 ACYL CARRIER (ACP) 2.
FT DOMAIN 3502 3570 ACYL CARRIER (ACP) 3.
FT DOMAIN 4544 4612 ACYL CARRIER (ACP) 4.
FT DOMAIN 6052 6129 ACYL CARRIER (ACP) 5.
FT BINDING 996 996 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 2028 2028 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 3532 3532 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 4574 4574 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 6082 6082 PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ SEQUENCE 6359 AA; 722923 MW; 82A273C546253074 CRC64;

Query Match 41.8%; Score 46; DB 1; Length 6359;
Best Local Similarity 50.0%; Pred. No. 1.9e+02;
Matches 11; Conservative 3; Mismatches 4; Indels 4; Gaps 1;

QY 1 VPSYFARHDSVEQDIIRDI 22
|||:| | | | | | | |
DB 3186 VPSFSF-----DSSVEDIFTTLI 3203

RESULT 10
YH75_ARCFU
ID YH75_ARCFU STANDARD; PRT; 330 AA.
AC O28499;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN AF1775.
GN AF1775.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=22334;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyte J.D.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).

CC -!- SIMILARITY: BELONGS TO THE ATZ/TRZ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000980; AAB89475.1; -.
CC TIGR; AF1775; -.
CC InterPro; IPR002604; ATZ_TRZ.
CC Pfam; PF01685; ATZ_TRZ; 1.
KW Hypothetical protein; Hydrolase; Complete proteome.
SQ SEQUENCE 330 AA; 37069 MW; 525AD9F7F35A6F61 CRC64;

Query Match 40.0%; Score 44; DB 1; Length 330;
Best Local Similarity 42.1%; Pred. No. 15;
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps

QY 3 SYSFARHDSVEQDIIRDI 21
|||:| | | | | | | |
DB 152 AYSARDIDLKMEEVREI 170

RESULT 11
RPOC_WEIHE
ID RPOC_WEIHE STANDARD; PRT; 1046 AA.
AC P96177;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE BETA' CHAIN (EC 2.7.7.6) (TRANSCRIPTASE
DE BETA' CHAIN) (RNA POLYMERASE BETA' SUBUNIT) (FRAGMENT).
GN RPOC.
OS Weissella hellenica.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Weissella.
OX NCBI_TaxID=46256;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=NCFB 2973;
RX MEDLINE=97016803; PubMed=8863429;
RA Morse R., Collins M.D., O'Hanlon K., Wallbanks S., Richardson P.T.;
RT "Analysis of the beta' subunit of DNA-dependent RNA polymerase does
RT not support the hypothesis inferred from 16S rRNA analysis that
RT Oenococcus oeni (formerly Leuconostoc oenos) is a tachytelic
RT (fast-evolving) bacterium."
RL Int. J. Syst. Bacteriol. 46:1004-1009(1996).
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC RNA(N).
CC -!- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
CC BETA' CHAIN.
CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X96470; CAA65322.1; -.
CC InterPro; IPR000722; RNA_pol_A.
CC Pfam; PF00623; RNA_pol_A; 1.
KW Transferase; DNA-directed RNA polymerase; Transcription.

```

Query Match 40.0%; Score 44; DB 1; Length 1058;  
Best Local Similarity 36.8%; Pred. No. 55;  
Matches 7; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

NCBI\_TaxID=10090;  
[1]  
RP SEQUENCE FROM N.A. (ISOFORMS M AND M1), AND VARIANTS MI AND MI-W.  
RN  
PP  
OX NCBI\_TaxID=10090;  
OX  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OC Eukaryota; metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
PP

CC TISSUE-Melanocyte;  
 CC MEDLINE=93345026; PubMed=8343963;  
 CC RA Hodgkinson C.A., Moore K.J., Nakayama A., Steingrimsson E.,  
 CC RA Copeland N.G., Jenkins N.A., Arnheiter H.;  
 CC RT "Mutations at the mouse microphthalmia locus are associated with  
 CC RT defects in a gene encoding a novel basic-helix-loop-helix-zipper  
 CC RT protein.";  
 CC RL Cell 74:395-404(1993).  
 CC RN [2]  
 CC RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND MUTAGENESIS.  
 CC RC STRAIN=129/SV; TISSUE=Heart;  
 CC RX MEDLINE=20253112; PubMed=10790403;  
 CC RA Hallsson J.H., Favor J., Hodgkinson C., Glaser T., Lamoreux M.L.,  
 CC RA Magnusdottir R., Gunnarsson G.J., Sweet H.O., Copeland N.G.,  
 CC RA Jenkins N.A., Steingrimsson E.;  
 CC RT "Genomic, transcriptional and mutational analysis of the mouse  
 CC RT microphthalmia locus.";  
 CC RL Genetics 155:291-300(2000).  
 CC RN [3]  
 CC RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS H AND M), AND VARIANTS.  
 CC RC STRAIN=C57BL/6; TISSUE=Heart, and Melanocyte;  
 CC RX MEDLINE=95179171; PubMed=7874168;  
 CC RA Steingrimsson E., Moore K.J., Lamoreux M.L., Ferre-D'Amare A.R.,  
 CC RA Burley S.K., Sanders Zmring D.C., Skow L.C., Hodgkinson C.A.,  
 CC RA Arnheiter H., Copeland N.G., Jenkins N.A.;  
 CC RT "Molecular basis of mouse microphthalmia (ml) mutations helps explain  
 CC RT their developmental and phenotypic consequences.";  
 CC RL Nat. Genet. 8:256-263(1994).  
 CC RN [4]  
 CC RP SEQUENCE OF 345-392 FROM N.A.  
 CC RC STRAIN=C57BL/6; TISSUE=Heart;  
 CC RX MEDLINE=94012591; PubMed=8407885;  
 CC RA Hughes M.J., Lingrel J.B., Krakowsky J.M., Anderson K.P.;  
 CC RT "A helix-loop-helix transcription factor-like gene is located at the  
 CC RT ml locus.";  
 CC RL J. Biol. Chem. 268:20687-20690(1993).  
 CC RN [5]  
 CC RP PARTIAL SEQUENCE FROM N.A. (ISOFORM A).  
 CC RX MEDLINE=98321192; PubMed=9647758;  
 CC RA Amai S., Fuse N., Yasumoto K.-I., Sato S., Yajima I., Yamamoto H.,  
 CC RA Uono T., Durlu Y.K., Tamai M., Takahashi K., Shibahara S.;  
 CC RT "Identification of a novel isoform of microphthalmia-associated  
 CC RT transcription factor that is enriched in retinal pigment epithelium.";  
 CC RL Biochem. Biophys. Res. Commun. 247:710-715(1998).  
 CC RN [6]  
 CC RP VARIANT MI-BW.  
 CC RX MEDLINE=99330550; PubMed=10400990;  
 CC RA Yajima I., Sato S., Kimura T., Yasumoto K.-I., Shibahara S.,  
 CC RA Goding C.R., Yamamoto H.;  
 CC RT "An L1 element intronic insertion in the black-eyed white (Mitfmi-bw)  
 CC RT gene: the loss of a single Mitf isoform responsible for the  
 CC RT pigmentary defect and inner ear deafness.";  
 CC RL Hum. Mol. Genet. 8:1431-1441(1999).  
 CC RN [7]  
 CC RP SUBCELLULAR LOCATION.  
 CC RX MEDLINE=96182124; PubMed=8622664;  
 CC RA Takebayashi K., Chida K., Tsukamoto I., Morii E., Munakata H.,  
 CC RA Arnheiter H., Kuroki T., Kitamura Y., Nomura S.;  
 CC RT "The recessive phenotype displayed by a dominant negative  
 CC RT microphthalmia-associated transcription factor mutant is a result of  
 CC RT impaired nucleation potential.";  
 CC RL Mol. Cell. Biol. 16:1203-1211(1996).  
 CC CC -!- FUNCTION: TRANSCRIPTION FACTOR FOR TYROSINASE AND TYROSINASE-  
 CC CC RELATED PROTEIN 1. BINDS TO A SYMMETRICAL DNA SEQUENCE (E-BOXES)  
 CC CC (5'-CAGGTG-3') FOUND IN THE TYROSINASE PROMOTER. PLAYS A CRITICAL  
 CC CC ROLE IN THE DIFFERENTIATION OF VARIOUS CELL TYPES AS NEURAL CREST-  
 CC CC DERIVED MELANOCYTES, MAST CELLS, OSTEOCLASTS AND OPTIC CUP-DERIVED  
 CC CC RETINAL PIGMENT EPITHELIUM.  
 CC CC -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC CC BHLH PROTEIN. BINDS DNA IN THE FORM OF HOMODIMER OR HETERODIMER  
 CC CC WITH EITHER TFEB, TFEB OR TFEC.  
 CC CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 CC CC -!- ALTERNATIVE PRODUCTS: 9 ISOFORMS; A (SHOWN HERE), A1, A2, H, H1,  
 CC CC H2, H3, M AND M1; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC CC -!- TISSUE SPECIFICITY: IN THE ADULT, EXPRESSED AT HIGH LEVELS IN THE  
 CC CC HEART, SKIN, SKELETAL MUSCLE, INTESTINE, STOMACH, KIDNEY, OVARY,  
 CC CC LUNG, SPLEEN AND BRAIN. IN THE EMBRYO, EXPRESSED IN DEVELOPING  
 CC CC EYE, EAR, SKIN AND HEART. ISOFORM M IS EXPRESSED IN MELANOCYTES  
 CC CC AND ALSO IN THE EMBRYONIC AND ADULT HEART WHILE ISOFORMS A AND H  
 CC CC ARE MORE WIDELY EXPRESSED.  
 CC CC -!- PTM: PHOSPHORYLATION AT SER-405 SIGNIFICANTLY ENHANCES THE ABILITY  
 CC CC TO BIND THE TYROSINASE PROMOTER (BY SIMILARITY).  
 CC CC -!- DISEASE: DEFECTS IN MITF ARE THE CAUSE OF MICROPTHALMIA (MI), A  
 CC CC CONDITION CHARACTERIZED BY LOSS OF PIGMENTATION; REDUCED EYE SIZE;  
 CC CC FAILURE OF SECONDARY BONE RESORPTION; REDUCED NUMBERS OF MAST  
 CC CC CELLS; EARLY ONSET OF DEAFNESS, AND WHICH GIVES RISE TO A NUMBER  
 CC CC OF DIFFERENT PHENOTYPES. AMONG THEM, MICROPTHALMIA-EYELESS WHITE  
 CC CC (MI-EW) HAS A NORMAL APPEARANCE AT THE HETEROZYGOUS STATE, BUT  
 CC CC SHOWS WHITE COAT; EYES ALMOST ABSENT AND EYELIDS NEVER OPEN AT  
 CC CC HOMOZYGOUSITY. MICROPTHALMIA-BLACK AND WHITE SPOTS (MI-BWS) IS  
 CC CC NORMAL AT HETEROZYGOUSITY, AND PRESENTS WHITE SPOTS AND BLACK EYES  
 CC CC AT HOMOZYGOUS STATE. MICROPTHALMIA-WHITE (MI-WH) HAS REDUCED COAT  
 CC CC COLOR AND EYE PIGMENTATION; SPOTS ON TOES, TAIL AND BELLY; INNER  
 CC CC EAR DEFECTS AT HETEROZYGOUSITY, AND AT HOMOZYGOUSITY SHOWS WHITE  
 CC CC COAT; EYES SMALL AND INNER IRIS SLIGHTLY PIGMENTED; SPINAL  
 CC CC GANGLIA, ADRENAL MEDULLA AND DERMIS SMALLER THAN NORMAL, AND INNER  
 CC CC EAR DEFECTS. MICROPTHALMIA-VITILIGO (MI-VI) HAS NORMAL PHENOTYPE  
 CC CC AT HETEROZYGOUSITY, BUT SHOWS GRADUAL DEPIGMENTATION OF COAT, SKIN  
 CC CC AND EYES; AND RETINAL DEGENERATION AT HOMOZYGOUSITY.  
 CC CC MICROPTHALMIA-SPOTTED (MI-SP) SHOWS NORMAL PHENOTYPE; AT  
 CC CC HOMOZYGOUSITY, HOWEVER, TYROSINASE ACTIVITY IN SKIN IS REDUCED.  
 CC CC MICROPTHALMIA-DEFECTIVE IRIS (MI-DI) HAS REDUCED RETINAL  
 CC CC PIGMENTATION AT HETEROZYGOUSITY AND SHOWS WHITE COAT; EYES OF  
 CC CC REDUCED SIZE AND POSSIBLE MILD OSTEOPOROSIS AT HOMOZYGOUSITY.  
 CC CC MICROPTHALMIA-CLOUDY EYED (MI-CE) HAS A NORMAL APPEARANCE AT THE  
 CC CC HETEROZYGOUS STATE, BUT SHOWS WHITE COAT; EYES OF REDUCED SIZE AND  
 CC CC UNPIGMENTED AT HOMOZYGOUSITY. MICROPTHALMIA-RED-EYED WHITE (MI-RW)  
 CC CC HAS A NORMAL APPEARANCE AT THE HOMOZYGOUS STATE, BUT SHOWS WHITE  
 CC CC COAT WITH ONE OR MORE PIGMENTED SPOTS AROUND THE HEAD/AND OR TAIL;  
 CC CC EYES ARE SMALL AND RED AT HETEROZYGOUSITY. MICROPTHALMIA-BLACK-  
 CC CC EYED WHITE (MI-BW) SHOWS A WHITE COAT BUT NORMAL SIZED EYES WHICH  
 CC CC REMAIN BLACK AT HOMOZYGOUSITY  
 CC CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.  
 CC CC -----  
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC CC the European Bioinformatics Institute. There are no restrictions on its  
 CC CC use by non-profit institutions as long as its content is in no way  
 CC CC modified and this statement is not removed. Usage by and for commercial  
 CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC CC -----  
 CC CC EMBL; Z23066; CAA80600.1;  
 CC CC EMBL; AF222344; AAF63466.1;  
 CC CC EMBL; AF222959; AAF81267.1;  
 CC CC EMBL; AF222950; AAF81267.1; JOINED.  
 CC CC EMBL; AF222951; AAF81267.1; JOINED.  
 CC CC EMBL; AF222953; AAF81267.1; JOINED.  
 CC CC EMBL; AF222954; AAF81267.1; JOINED.  
 CC CC EMBL; AF222955; AAF81267.1; JOINED.  
 CC CC EMBL; AF222956; AAF81267.1; JOINED.  
 CC CC EMBL; AF222957; AAF81267.1; JOINED.  
 CC CC EMBL; AF222958; AAF81267.1; JOINED.  
 CC CC EMBL; AF222959; AAF81269.1;  
 CC CC EMBL; AF222950; AAF81269.1; JOINED.  
 CC CC EMBL; AF222951; AAF81269.1; JOINED.  
 CC CC EMBL; AF222956; AAF81269.1; JOINED.  
 CC CC EMBL; AF222958; AAF81269.1; JOINED.  
 CC CC EMBL; AF222959; AAF81270.1;  
 CC CC EMBL; AF222950; AAF81270.1; JOINED.  
 CC CC EMBL; AF222951; AAF81270.1; JOINED.  
 CC CC EMBL; AF222953; AAF81270.1; JOINED.  
 CC CC EMBL; AF222954; AAF81270.1; JOINED.  
 CC CC EMBL; AF222955; AAF81270.1; JOINED.  
 CC CC EMBL; AF222956; AAF81270.1; JOINED.



DR InterPro; IPR000362; Fumarate\_lyase.  
DR Pfam; PF00206; lyase\_1; 1.  
DR PROSITE; PS00163; FUMARATE\_LYASES; 1.  
KW Lyase; Tricarboxylic acid cycle; Iron-sulfur; 4Fe-4S;  
KW Complete proteome.  
FT INIT\_MET 0  
FT METAL 317 317 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
FT ACT\_SITE 396 396 POTENTIAL.  
FT BINDING 462 462 CARBOXYL GROUP (POTENTIAL).  
SQ SEQUENCE 547 AA; 60167 MW; F9827451050334D8 CRC64;

Query Match 39.1%; Score 43; DB 1; Length 547;  
Best Local Similarity 44.4%; Pred. No. 38;  
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 3 SYSFARAHDSVQDIIRD 20  
I: ||:|:|:|:  
Db 58 SFMLRPAHQVADILRD 75

Search completed: March 27, 2002, 14:27:09  
Job time: 1651 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:01:24 ; Search time 102.51 seconds  
(without alignments)  
16.348 Million cell updates/sec

Title: US-09-290-049a-18

Perfect score: 110

Sequence: 1 VPNVVFIRAHDSVQTRIAKII 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 68:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	110	100.0	1365	2 A41483	glucosyltransferase
2	90	81.8	1431	2 A45866	dextranase (EC 2.4.1.5)
3	90	81.8	1508	2 T31098	probable dextranase
4	86	78.2	1290	2 JC5473	dextranase (EC 2.4.1.5)
5	80	72.7	1475	2 B33135	gtfB protein precursor
6	77	70.0	1577	2 T30858	glucosyltransferase
7	76	69.1	1375	2 JT0345	dextranase (EC 2.4.1.5)
8	75	68.2	1599	2 S22737	glucosyltransferase
9	74	67.3	1449	2 T30857	glucosyltransferase
10	74	67.3	1449	2 T30857	glucosyltransferase
11	73	66.4	1518	2 A44811	glucosyltransferase
12	69	62.7	1592	2 A38175	glucosyltransferase
13	49	44.5	602	2 A71256	aspartate--tRNA ligase
14	48	43.6	492	2 C72417	sugar kinase, FGGI
15	47	42.7	506	2 E70155	aspartate--tRNA ligase
16	47	42.7	557	2 S73434	aspartate--tRNA ligase
17	45	40.9	654	1 BVBP1	CBP1 protein - yea
18	44	40.0	583	2 T04531	nine-cis-epoxycarotene dioxygenase
19	43	39.1	457	2 H71553	probable biotin carboxyl carrier protein
20	43	39.1	457	2 D81708	acetyl-coenzyme A carboxylase
21	43	39.1	765	2 S76795	hypothetical protein
22	43	39.1	1070	2 S75712	cellulase (EC 3.2.1.4)
23	42.5	38.6	92	2 E47754	ycaA protein homolog
24	42.5	38.6	319	2 A86777	conserved hypothetical protein
25	42	38.2	144	2 H83933	heat shock protein
26	42	38.2	346	2 E70715	hypothetical protein
27	41	37.3	160	2 S73763	transcription elongation factor
28	41	37.3	161	2 B64231	transcription elongation factor
29	41	37.3	270	2 S44952	ImbE protein - Str

30 41 37.3 488 2 B82798 virulence-associated  
31 41 37.3 560 1 JC4795 plasma hyaluronan-  
32 41 37.3 688 2 T21641 hypothetical prote  
33 41 37.3 805 2 T21957 hypothetical prote  
34 41 37.3 815 2 T05555 DNA polymerase III  
35 41 37.3 961 1 P1BVA RNA Ia protein - b  
36 41 37.3 1428 2 S62419 hypothetical prote  
37 41 37.3 13055 2 T16580 hypothetical prote  
38 40 36.4 279 2 T27854 hypothetical prote  
39 40 36.4 354 2 S65887 (A+T)-stretch-bind  
40 40 36.4 594 2 T38114 pyruvate decarboxy  
41 40 36.4 605 2 T43191 probable pyruvate  
42 40 36.4 652 2 T20549 hypothetical prote  
43 40 36.4 933 2 G70166 probable zinc prot  
44 40 36.4 943 2 B48474 glycoprotein B - f  
45 40 36.4 948 2 A56602 glycoprotein B hom

#### ALIGNMENTS

RESULT 1

A41483

glucosyltransferase (EC 2.4.1.5) gtfS precursor - Streptococcus sobrinus

C:Species: Streptococcus sobrinus

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 15-Oct-1999

C:Accession: A41483

R:Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.

Infect. Immun. 58, 2452-2458, 1990

A:Title: Analysis of the Streptococcus downei gtfS gene, which specifies a glucosyltr

A:Reference number: A41483; MUID:90316665

A:Accession: A41483

A:Molecule type: DNA

A:Residues: 1-1365 <GIL>

A:Cross-references: GB:M30943; NID:G153652; PIDN:AAA26898.1; PID:G153653

C:Genetics:

A:Gene: gtfS

C:Superfamily: cpl repeat homology

C:Keywords: glucosyltransferase; hexosyltransferase

Query Match 100.0%; Score 110; DB 2; Length 1365;  
Best Local Similarity 100.0%; Pred. No. 7.2e-10;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPNVVFIRAHDSVQTRIAKII 22

DB 537 VPNVVFIRAHDSVQTRIAKII 558

RESULT 2

A45866

dextranase (EC 2.4.1.5) precursor - Streptococcus mutans

C:Species: Streptococcus mutans

C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999

C:Accession: A45866

R:Honda, O.; Kato, C.; Kuramitsu, H.K.

J. Gen. Microbiol. 136, 2099-2105, 1990

A:Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the gluco

A:Reference number: A45866; MUID:91100958

A:Accession: A45866

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1431 <HON>

A:Cross-references: GB:M29296

C:Superfamily: cpl repeat homology

C:Keywords: glucosyltransferase; hexosyltransferase

F:181-201/Domain: cpl repeat homology <CP1>

F:1127-1146/Domain: cpl repeat homology <CP2>

F:1192-1211/Domain: cpl repeat homology <CP3>

F:1257-1276/Domain: cpl repeat homology <CP4>

F:1277-1297/Domain: cpl repeat homology <CP5>

F:1321-1340/Domain: cpl repeat homology <CP8>

F:1341-1361/Domain: cpl repeat homology <CP6>  
F:1385-1404/Domain: cpl repeat homology <CP7>

Query Match 81.8%; Score 90; DB 2; Length 1431;  
Best Local Similarity 90.0%; Pred. No. 1.7e-06;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NYVVFIRAHDSVQVTRIAKII 22  
DB 576 NYIFVIRAHDSVQVTRIAKII 595

RESULT 3

T31098  
probable dextranucrase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides  
C:Species: Leuconostoc mesenteroides  
C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000  
C:Accession: T31098  
R:Monchois, V.; Renaud-Simeon, M.; Monsan, P.; Willemot, R.M.  
FEMS Microbiol. Lett. 159, 307-315, 1998  
A:Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase (DS)  
A:Reference number: 220981; MUID:98164374  
A:Accession: T31098  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1508 <MON>  
A:Cross-references: EMBL:AF030129; NID:g2766611; PID:g2766612; PIDN:ANB95453.1  
A:Experimental source: strain NRRL B-1299  
C:Genetics:  
A:Gene: dsrB  
C:Function:  
A:Description: produces dextran composed only of alpha(1-6) glucosidic bonds  
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 81.8%; Score 90; DB 2; Length 1508;  
Best Local Similarity 77.3%; Pred. No. 1.8e-06;  
Matches 17; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPNVFIRAHDSVQVTRIAKII 22  
DB 634 IPNYSFVIRAHDSVQVTRIAKII 655

RESULT 4

T30873  
dextranucrase (EC 2.4.1.5) - Leuconostoc mesenteroides  
C:Species: Leuconostoc mesenteroides  
C>Date: 07-Jul-1997 #sequence\_revision 29-Aug-1997 #text\_change 29-Aug-1997  
C:Accession: JC5473  
R:Monchois, V.; Willemot, R.M.; Renaud-Simeon, M.; Croux, C.; Monsan, P.  
Gene 182, 23-32, 1996  
A:Title: Cloning and sequencing of a gene coding for a novel dextranucrase from Leucon  
A:Reference number: JC5473; MUID:97136686  
A:Accession: JC5473  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1290 <MON>  
A:Cross-references: GB:U38181  
C:Comment: This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose ont  
C:Genetics:  
A:Gene: dsrA  
C:Keywords: glycosyltransferase; hexosyltransferase  
F:78-870/Domain: catalytic #status predicted <CAT>  
F:922-1290/Domain: glucan-binding #status predicted <GCB>

Query Match 78.2%; Score 86; DB 2; Length 1290;  
Best Local Similarity 85.7%; Pred. No. 6.8e-06;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 PNVVFIRAHDSVQVTRIAKII 22

DB 388 PNYSFIRAHDSVQVTRIAKII 408

RESULT 5

B33135  
gtfB protein precursor - Streptococcus mutans  
C:Species: Streptococcus mutans  
C>Date: 23-Oct-1990 #sequence\_revision 23-Oct-1990 #text\_change 15-Oct-1999  
C:Accession: B33135; A33128  
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
J. Bacteriol. 169, 4263-4270, 1987  
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.  
A:Reference number: A33135; MUID:87308013  
A:Accession: B33135  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1475 <SHI>  
A:Cross-references: GB:M17361; NID:g153639; PIDN:AAA88588.1; PID:g153640  
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
submitted to the Protein Sequence Database, September 1990  
A:Reference number: A33128  
A:Accession: A33128  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-171,173-641,'N',643-1475 <SH2>  
A:Experimental source: strain GS-5  
C:Superfamily: cpl repeat homology  
F:1096-1115/Domain: cpl repeat homology <CP1>  
F:1224-1243/Domain: cpl repeat homology <CP2>  
F:1289-1308/Domain: cpl repeat homology <CP3>  
F:1354-1373/Domain: cpl repeat homology <CP4>  
F:1419-1438/Domain: cpl repeat homology <CP5>

Query Match 72.7%; Score 80; DB 2; Length 1475;  
Best Local Similarity 77.3%; Pred. No. 8e-05;  
Matches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPNVFIRAHDSVQVTRIAKII 22  
DB 552 VPSYFIRAHDSVQVTRIAKII 573

RESULT 6

T30858  
glucosyltransferase - Streptococcus salivarius  
C:Species: Streptococcus salivarius  
C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30858  
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.  
Infect. Immun. 63, 609-621, 1995  
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for  
A:Reference number: 220909; MUID:95122197  
A:Accession: T30858  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1577 <SIM>  
A:Cross-references: EMBL:L35928; NID:g662380; PID:g662381; PIDN:AAC41413.1  
C:Genetics:  
A:Gene: gtfm

Query Match 70.0%; Score 77; DB 2; Length 1577;  
Best Local Similarity 70.0%; Pred. No. 0.00027;  
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 NYVVFIRAHDSVQVTRIAKII 22

DB 661 NYIFVIRAHDSVQVTRIAKII 680

RESULT 7



JT0345  
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)  
N:Alternate names: sucrose 6-glucosyltransferase  
C:Species: Streptococcus mutans  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 24-Sep-1999  
C:Accession: JT0345; C33135  
R:Ueda, S.; Shiroza, T.; Kuramitsu, H.K.  
Gene 69, 101-109, 1988  
A:Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.  
A:Reference number: JT0345; MUID:89137980  
A:Accession: JT0345  
A:Molecule type: DNA  
A:Residues: 1-1375 <OED>  
A:Experimental source: GS-5  
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
J. Bacteriol. 169, 4263-4270, 1987  
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.  
A:Reference number: A33135; MUID:87308013  
A:Accession: C33135  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-349 <SHI>  
A:Cross-references: GB:M17361  
C:Genetics:  
A:Gene: gtfC  
C:Function:  
A:Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans  
C:Superfamily: cpl repeat homology  
C:Keywords: duplication; glucosyltransferase; hexosyltransferase  
F:1-34/Domain: signal sequence #status predicted <SIG>  
F:35-1375/Product: glucosyltransferase #status predicted <MAT>  
F:1126-1145/Domain: cpl repeat homology <CPL>  
F:1253-1272/Domain: cpl repeat homology <CP2>  
F:1318-1337/Domain: cpl repeat homology <CP3>

Query Match 69.1%; Score 76; DB 2; Length 1375;  
Best Local Similarity 72.7%; Pred. No. 0.00034;  
Matches 16; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 VPNTVFIRAHDSVQTRIAKII 22  
||| ||||| ||||| |||  
Db 578 VPSYFIRAHDSVQDLIRNII 599

RESULT 8  
S22737  
glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius  
C:Species: Streptococcus salivarius  
C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 21-Jan-2000  
C:Accession: S22737; S28810; B44811; S22727  
R:Jacques, N.  
submitted to the EMBL Data Library, March 1992  
A:Reference number: S22726  
A:Accession: S22737  
A:Molecule type: DNA  
A:Residues: 1-1599 <YAC>  
A:Cross-references: EMBL:Z11872; NID:g47530; PIDN:CAA77898.1; PID:g47531  
A:Experimental source: ATCC 25975  
R:Giffard, P.M.; Simpson, C.L.; Millward, C.P.; Jacques, N.A.  
J. Gen. Microbiol. 137, 2577-2593, 1991  
A:Title: Molecular characterization of a cluster of at least two glucosyltransferase genes  
A:Reference number: A44811; MUID:92148377  
A:Accession: S28810  
A:Molecule type: DNA  
A:Residues: 1-51 <GIF>  
A:Cross-references: EMBL:Z11873  
C:Genetics:  
A:Gene: gtfC  
C:Superfamily: cpl repeat homology  
C:Keywords: glucosyltransferase; hexosyltransferase  
F:1456-1475/Domain: cpl repeat homology <CPR>

Query Match 68.2%; Score 75; DB 2; Length 1599;  
Best Local Similarity 78.9%; Pred. No. 0.0006;  
Matches 15; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 YVFIRAHDSVQTRIAKII 22  
||| ||||| ||||| |||  
Db 575 YLFVRAHDSVQTVIADII 593

RESULT 9  
T30857  
glucosyltransferase - Streptococcus salivarius  
C:Species: Streptococcus salivarius  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30857  
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.  
Infect. Immun. 63, 609-621, 1995  
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for  
A:Reference number: Z20909; MUID:95122197  
A:Accession: T30857  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1449 <SIM>  
A:Cross-references: EMBL:L35495; NID:g662378; PID:g662379; PIDN:AAC41412.1  
C:Genetics:  
A:Gene: gtfL

Query Match 67.3%; Score 74; DB 2; Length 1449;  
Best Local Similarity 70.0%; Pred. No. 0.00078;  
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 NYVFIRAHDSVQTRIAKII 22  
||| ||||| ||||| |||  
Db 609 NYAFVRAHDSVQSIIGQII 628

RESULT 10  
T30552  
glucosyltransferase N - Streptococcus salivarius (fragment)  
C:Species: Streptococcus salivarius  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30552  
R:Jaffe, R.I.  
submitted to the EMBL Data Library, February 1998  
A:Description: Streptococcus salivarius VI477 gtfN.  
A:Reference number: Z20854  
A:Accession: T30552  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1449 <JAF>  
A:Cross-references: EMBL:AF049609; NID:g2935545; PID:g2935546; PIDN:AAC05156.1  
C:Genetics:  
A:Gene: gtfN

Query Match 67.3%; Score 74; DB 2; Length 1449;  
Best Local Similarity 70.0%; Pred. No. 0.00078;  
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 NYVFIRAHDSVQTRIAKII 22  
||| ||||| ||||| |||  
Db 609 NYAFVRAHDSVQSIIGQII 628

RESULT 11  
A44811  
glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius  
C:Species: Streptococcus salivarius  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 15-Oct-1999  
C:Accession: A44811; S22726; S28809  
R:Giffard, P.M.; Simpson, C.L.; Millward, C.P.; Jacques, N.A.

J. Gen. Microbiol. 137, 2577-2593, 1991  
 A:Title: Molecular characterization of a cluster of at least two glucosyltransferase genes  
 A:Reference number: A44811; MUID:92148377  
 A:Accession: A44811  
 A:Molecule type: DNA  
 A:Residues: 1-1518 <GIF>  
 A:Cross-references: EMBL:211873; NID:g47526; PIDN:CAA77900.1; PID:g47527  
 A:Note: sequence extracted from NCBI backbone (NCBIN:81050, NCBIP:81052)  
 C:Genetics:  
 A:Gene: gtfJ  
 C:Superfamily: cpl repeat homology  
 C:Keywords: glucosyltransferase; hexosyltransferase  
 F:1307-1326/Domain: cpl repeat homology <CP4>

Query Match 66.4%; Score 73; DB 2; Length 1518;  
 Best Local Similarity 75.0%; Pred. No. 0.0012;  
 Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

3 NYVFIRAHDSVQTRIAKII 22  
 |||||  
 Db 604 NYVFIRAHDSVQTRIAKII 623  
 |||||

RESULT 12  
 A38175  
 glucosyltransferase precursor - Streptococcus sobrinus  
 C:Species: Streptococcus sobrinus  
 C:Date: 28-Aug-1992 #sequence\_revision 28-Aug-1992 #text\_change 15-Oct-1999  
 C:Accession: A38175  
 R:Abu, H.; Mtsamura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.  
 J. Bacteriol. 173, 989-996, 1991  
 A:Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus  
 A:Reference number: A38175; MUID:91123227  
 A:Accession: A38175  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1592 <ABO>  
 C:Cross-references: GB:D90213; NID:g217032; PIDN:BAA14241.1; PID:d1014946; PID:g217033  
 C:Superfamily: cpl repeat homology  
 F:1093-1112/Domain: cpl repeat homology <CP1>  
 F:1222-1241/Domain: cpl repeat homology <CP2>  
 F:1287-1306/Domain: cpl repeat homology <CP3>  
 F:1330-1351/Domain: cpl repeat homology <CP4>  
 F:1352-1371/Domain: cpl repeat homology <CP5>  
 F:1402-1420/Domain: cpl repeat homology <CP6>  
 F:1465-1484/Domain: cpl repeat homology <CP7>  
 F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match 62.7%; Score 69; DB 2; Length 1592;  
 Best Local Similarity 68.2%; Pred. No. 0.006;  
 Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 VPNVFIRAHDSVQTRIAKII 22  
 |||||  
 Db 548 VPSYFARAHDSVQDIIRDII 569  
 |||||

RESULT 13  
 A71256  
 aspartate--trna ligase (EC 6.1.1.12) - syphilis spirochete  
 N:Alternate names: aspartyl--trna synthetase TP0985  
 C:Species: Treponema pallidum subsp. pallidum [syphilis spirochete]  
 C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 16-Jul-1999  
 C:Accession: A71256  
 R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin  
 rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo  
 they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
 Science 281, 375-388, 1998  
 A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
 A:Reference number: A71250; MUID:98332770  
 A:Accession: A71256

A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-602 <COL>  
 A:Cross-references: GB:AE001266; GB:AE000520; NID:g3323309; PIDN:AAC65942.1; PID:g332  
 A:Experimental source: strain Nichols  
 C:Genetics:  
 A:Gene: TP0985  
 C:Function:  
 A:Description: activates amino acid and transfers it to specific trna molecule  
 A:Pathway: protein biosynthesis  
 C:Superfamily: lysine--trna ligase  
 C:Keywords: aminoacyl--trna synthetase; ATP; ligase; protein biosynthesis

Query Match 44.5%; Score 49; DB 2; Length 602;  
 Best Local Similarity 56.2%; Pred. No. 4.2;  
 Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 7 IRAHDSVQTRIAKII 22  
 |||||  
 Db 505 IRHDTOLQKRFKIV 520  
 |||||

RESULT 14  
 C72417  
 sugar kinase, FGGY family - Thermotoga maritima (strain MSB8)  
 C:Species: Thermotoga maritima  
 C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
 C:Accession: C72417  
 R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic  
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,  
 C.M.  
 Nature 399, 323-329, 1999  
 A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome  
 A:Reference number: A72200; MUID:99287316  
 A:Accession: C72417  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-492 <ARN>  
 A:Cross-references: GB:AE001697; GB:AE000512; NID:g4980597; PIDN:AAD35210.1; PID:g498  
 A:Experimental source: strain MSB8  
 C:Genetics:  
 A:Gene: TM0116  
 C:Superfamily: xylulokinase

Query Match 43.6%; Score 48; DB 2; Length 492;  
 Best Local Similarity 36.4%; Pred. No. 4.8;  
 Matches 8; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

Qy 1 VPNVFIRAHDSVQTRIAKII 22  
 |||||  
 Db 136 LPKILWIRKHEPEYGRKSKIM 157  
 |||||

RESULT 15  
 E70155  
 aspartate--trna ligase (EC 6.1.1.12) asps - Lyme disease spirochete  
 N:Alternate names: aspartyl--trna synthetase  
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
 C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 24-Nov-1999  
 C:Accession: E70155  
 R:Fraser, C.M.; Castjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh  
 son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu  
 ; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
 Nature 390, 580-586, 1997  
 A:Authors: Smith, H.O.; Venter, J.C.  
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.  
 A:Reference number: A70100; MUID:98065943  
 A:Accession: E70155  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-506 <KLE>

A;Cross-references: GB:AE001149; GB:AE000783; NID:g2688348; PIDN:AAB91506.1; PID:g268834  
A;Experimental source: strain B31  
C;Superfamily: lysine--tRNA ligase  
C;Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 42.7%; Score 47; DB 2; Length 506;  
Best Local Similarity 62.5%; Pred. No. 7.3;  
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 7 IRAHDSEVQTRIAKII 22  
|||:|:|:|:|:|  
Db 406 IRIHNKELOQRIPKII 421

Search completed: March 27, 2002, 14:01:24  
Job time: 486 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:27:08 ; Search time 53.4 Seconds  
(without alignments)  
15.105 Million cell updates/sec

Title: US-09-290-049a-18

Perfect score: 110

Sequence: 1 VPNVYFIRAHDSVQTRIAKII 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Maximum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	100.0	1365	1	GTFS_STRDO
2	90	81.8	1462	1	GTFD_STRMU
3	76	69.1	1375	1	GTFC_STRMU
4	75	68.2	1476	1	GTFB_STRMU
5	70	63.6	1597	1	GTFI_STRDO
6	69	62.7	1592	1	GTFL_STRDO
7	49	44.5	602	1	SYD_TREPA
8	47	42.7	336	1	RA51_DROME
9	47	42.7	537	1	SYD_MYCPN
10	47	42.7	586	1	SYD_BORBU
11	45	40.9	654	1	CBP1_YEAST
12	42	38.2	346	1	Y943_MYCTU
13	41	37.3	160	1	GRE4_MYCPN
14	41	37.3	161	1	GRE4_MYCGE
15	41	37.3	223	1	KAD4_MOUSE
16	41	37.3	223	1	KAD4_RAT
17	41	37.3	584	1	SYD_BUCAP
18	41	37.3	737	1	SKN1_CANAL
19	41	37.3	961	1	VLA_BMV
20	41	37.3	1428	1	YA84_SCHPO
21	40	36.4	205	1	ADEN_ADEG8
22	40	36.4	594	1	DCP2_SCHPO
23	40	36.4	2261	1	RRPL_MUMPM
24	39.5	35.9	366	1	RRPO_REOVJ
25	39.5	35.9	445	1	DHA3_BACSU
26	39	35.5	206	1	ADEN_ADECU
27	39	35.5	236	1	PHOU_XYLFA
28	39	35.5	297	1	LE33_CAEEL
29	39	35.5	580	1	SYD_THETH
30	39	35.5	606	1	SP2_HUMAN
31	39	35.5	2183	1	RRPL_RINDR
32	39	35.5	2184	1	RRPL_CDVO
33	39	35.5	2672	1	GCN1_YEAST

RESULT 1

ID	GTFS_STRDO	STANDARD;	PRT;	1365 AA.
AC	P29336;			
DT	01-DEC-1992 (Rel. 24, Created)			
DT	01-DEC-1992 (Rel. 24, Last sequence update)			
DT	01-APR-1993 (Rel. 25, Last annotation update)			
DE	GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)			
DE	(SUCROSE 6-GLUCOSYLTRANSFERASE).			
GN	GTFS.			
OS	Streptococcus downei (Streptococcus sobrinus).			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;			
OC	Streptococcus.			
OX	NCBI_TaxID=1317;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=MFE28;			
RC	MEDLINE=90316665; PubMed=2142479;			
RA	Gilmore K.S., Russell R.R., Ferretti J.J.;			
RA	"Analysis of the Streptococcus downei gtfS gene, which specifies a			
RT	glucosyltransferase that synthesizes soluble glucans.";			
RL	Infect. Immun. 58:2452-2458(1990).			
CC	-!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT			
CC	TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE			
CC	OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE			
CC	AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.			
CC	-!- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =			
CC	D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).			
CC	-!- ENZYME REGULATION: GLUCAN SYNTHESIS BY GTF-S IS INDEPENDENT OF			
CC	PRIMER GLUCAN UNLIKE GTF-1.			
CC	-!- DISEASE: DENTAL CARIES.			
CC	-!- MISCELLANEOUS: GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA			
CC	1,6-GLUCOSE).			
CC	-!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-			
CC	BINDING PROTEIN FROM S. MUTANS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; M30943; AAA26898.1; -.			
DR	PIR; A41483; A41483.			
DR	InterPro; IPR002479; CW_binding.			
DR	InterPro; IPR003318; Glyco_hydro_70.			
DR	Pfam; PF01473; CW_binding_1; 10.			
DR	Pfam; PF02324; Glyco_hydro_70; 1.			
KW	Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.			
FT	SIGNAL 1 36 OR 37 (POTENTIAL).			
FT	CHAIN 37 1365 GLUCOSYLTRANSFERASE-S.			
FT	DOMAIN 37 1050 CATALYTIC (APPROXIMATE).			
FT	DOMAIN 1083 1365 GLUCAN-BINDING (APPROXIMATE).			
FT	DOMAIN 1083 1365 4.5 X TANDEM REPEATS.			

34	38.5	35.0	366	1	RRPO_REOVL
35	38.5	35.0	366	1	RRPO_REOVL
36	38.5	35.0	709	1	VM21_REOVL
37	38	34.5	289	1	DHPS_SYNY3
38	38	34.5	355	1	CYSA_SYNY3
39	38	34.5	368	1	CCR3_HUMAN
40	38	34.5	386	1	SUC_RICPR
41	38	34.5	446	1	SYG_MYCGE
42	38	34.5	490	1	IMDH_AQUAE
43	38	34.5	501	1	XYLB_LACIA
44	38	34.5	641	1	PRIM_UREPA
45	38	34.5	1138	1	DPS2_YEAST

ALIGNMENTS



```

RESULT 3
ID GTFC_STRMU STANDARD; PRT: 1375 AA.
AC P13470; P05427;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI PRECURSOR (EC 2.4.1.5) (GTF-SI)
DE (DEXTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTFC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=89137980; PubMed=2976010;
RA Ueda S., Shiroza T., Kuramitsu H.K.;
RT "Sequence analysis of the gtfc gene from Streptococcus mutans GS-5.";
F Gene 69:101-109(1988).
[2]
K SEQUENCE OF 1-349 FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfc gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
CC -I- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -I- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) =
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).
CC -I- SUBCELLULAR LOCATION: SECRETED.
CC -I- DISEASE: DENTAL CARIES.
CC -I- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -I- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S.MUTANS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M22054; AAA88592.1; -
DR EMBL; M17361; AAA88589.1; -
DR PIR; JT0345; JT0345.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 7.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1 34
FT CHAIN 35 1375 GLUCOSYLTRANSFERASE-SI.
FT DOMAIN 35 1050 CATALYTIC (APPROXIMATE).
FT DOMAIN 1126 1375 GLUCAN-BINDING (APPROXIMATE).
FT DOMAIN 1126 1375 2.4 A, 1 C AND 1 AC REPEATS.
FT REPEAT 1126 1159 A REPEAT.
FT REPEAT 1169 1200 A REPEAT.
FT REPEAT 1227 1238 C REPEAT.
FT REPEAT 1253 1303 AC REPEAT.
FT REPEAT 1318 1330 A REPEAT (INCOMPLETE).
SQ SEQUENCE 1375 AA; 153022 MW; D4B80CBEE0AAACE13 CRC64;

Query Match 69.1%; Score 76; DB 1; Length 1375;
Best Local Similarity 72.7%; Pred. No. 0.00018;
Matches 16; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 VPNYVFIRAHDSVQTRIAKII 22
DB 578 VPSYFIRAHDSVQDLIRNII 599

RESULT 4
ID GTFB_STRMU STANDARD; PRT: 1476 AA.
AC P08987; O69381; O69384; O69387; O69390; O69396;
DT 01-NOV-1988 (Rel. 09, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTFB.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfb gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
RN [2]
K SEQUENCE FROM N.A.
RC STRAIN=MT4239, MT4245, MT4251, MT4467, AND MT8148;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
CC -I- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -I- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) =
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).
CC -I- SUBCELLULAR LOCATION: SECRETED.
CC -I- DISEASE: DENTAL CARIES.
CC -I- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -I- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S.MUTANS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M17361; AAA88588.1; -
DR EMBL; D88651; BAA26101.1; -
DR EMBL; D88654; BAA26105.1; -
DR EMBL; D88657; BAA26109.1; -
DR EMBL; D88660; BAA26113.1; -
DR EMBL; D89977; BAA26119.1; -
DR PIR; B33135; B33135.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 13.
DR Pfam; PF02324; Glyco_hydro_70; 1.

```





```

Query Match      62.7%;   Score 69;   DB 1;   Length 1592;
Best Local Similarity 68.2%;   Pred. No. 0.003;
Matches 15;   Conservative 1;   Mismatches 6;   Indels 0;   Gaps 0;

QY      1 VPNYVFIRAHDSVOTRIAKII 22
      ||| | ||||| | ||

```



```
DR InterPro; IPR002309; tRNA-synt.2.
DR InterPro; IPR002312; tRNA-synt_2.1.
DR Pfam; PF00152; tRNA-synt_2.1.
DR Pfam; PF01336; tRNA-anti; 1.
DR PRINTS; PR01042; TRNASYNTHASP.
DR PROSITE; PS00179; AA_TRNA_LIGASE_II.1; FALSE_NEG.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II.2; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 557 AA; 64114 MW; 65EBE73A85F8A424 CRC64;

Query Match 42.7%; Score 47; DB 1; Length 557;
Best Local Similarity 75.0%; Pred. No. 4.1;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 7 IRAHSEVQTRI 18
II I I I I I I I
Db 463 IRIHDEVQTRL 474

LT 10
SYD_BORBU
ID SYD_BORBU STANDARD; PRT; 586 AA.
AC 051402;
DT 15-DEC-1998 (Rel. 37, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ASPARYL-TRNA SYNTHETASE (EC 6.1.1.12) (ASPARTATE--TRNA LIGASE)
DE (ASPRS).
GN ASPS OR BB0446.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Peterson J., Tomb J.F., Fleischmann R.D., Richardson D.,
RA van Vugt R., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA Utterback T., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Garland S., Fujii C., Cotton M.D., Artiach P., Bowman C.,
RA Smith H.O., Venter J.C.; Cotton M.D., Horst K., Roberts K., Hatch B.,
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT burgdorferi."
BL Nature 390:580-586(1997).
CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + TRNA(ASP) = AMP +
CC PYROPHOSPHATE + L-ASPARTYL-TRNA(ASP).
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001149; AAB91506.1; ALT_INIT.
CC TIGR; BB0446;
DR InterPro; IPR002106; AA_trna_ligase_II.
DR InterPro; IPR002309; tRNA-synt_2.
DR InterPro; IPR002312; tRNA-synt_2.
DR InterPro; IPR002313; tRNA-synt_lys_2.
DR Pfam; PF00152; tRNA-synt_2; 2.
DR Pfam; PF01336; tRNA-anti; 1.
DR PRINTS; PR00982; TRNASYNTHLSP.
DR PRINTS; PR01042; TRNASYNTHASP.

DR PROSITE; PS00179; AA_TRNA_LIGASE_II.1; 1.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II.2; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 586 AA; 68343 MW; B7BAE0EC44BE16EB CRC64;

Query Match 42.7%; Score 47; DB 1; Length 586;
Best Local Similarity 62.5%; Pred. No. 4.3;
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Oy 7 IRAHSEVQTRI 22
II I I I I I I I
Db 486 IRIHKEAQRIKFI 501

RESULT 11
CBPI_YEAST
ID CBPI_YEAST STANDARD; PRT; 654 AA.
AC P07252;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CYTOCHROME B PRE-MRNA PROCESSING PROTEIN 1.
GN CBPI OR YJL209W OR J0242 OR HRA654.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84185566; PubMed=6325407;
RA Dieckmann C.L., Homison G., Tzagoloff A.;
RT "Assembly of the mitochondrial membrane system. Nucleotide sequence
RT of a yeast nuclear gene (CBPI) involved in 5' end processing of
RT cytochrome b pre-mRNA."
RL J. Biol. Chem 259:4732-4738(1984).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=95242842; PubMed=7725802;
RA Vandenbol M., Durand P., Bolle P.-A., Dion C., Portetelle D.,
RA Hilger F.;
RT "Sequence analysis of a 40.2 kb DNA fragment located near the left
RT telomere of yeast chromosome X."
RL Yeast 10:1657-1662(1994).
RN [3]
RP SEQUENCE OF 589-654 FROM N.A.
RX MEDLINE=90014786; PubMed=2552292;
RA Liu Y., Dieckmann C.L.;
RT "Overproduction of yeast viruslike particles by strains deficient in
RT a mitochondrial nuclease."
RL Mol. Cell. Biol. 9:3323-3331(1989).
CC -1- FUNCTION: RESPONSIBLE FOR CONFERRING A STABLE 5' END ON CYTOCHROME
CC B MRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; K02647; AAA34474.1;
CC EMBL; Z34098; CAA84002.1;
CC EMBL; Z49484; CAA89506.1;
CC EMBL; M28067; AAA34456.1;
CC PIR; S05829; BVBYPI.
CC PIR; S45164; S45164.
CC SGD; S0003745; CBPI.
CC mRNA processing.
CC SEQUENCE 654 AA; 76171 MW; 2453B03280E1C44D CRC64;
SQ
```

```

Query Match          40.9%; Score 45; DB 1; Length 654;
Best Local Similarity 47.4%; Pred. No. 10;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 3 NVVFIRAHDSVQTRIARI 21
Db 300 HVSISAHDSRIIYKVKFI 318
      :||| | ||| : :|||

RESULT 12
Y943_MYCTU
ID Y943_MYCTU STANDARD; PRT; 346 AA.
AC P71566;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 37.2 KDA PROTEIN RV0943C.
RV0943C OR MT0969 OR MTCY10D7.31.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch K., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains.";
Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; Z79700; CAB01978.1; -.
DR EMBL; AE006982; AAK45217.1; ALT_INIT.
DR TIGR; MT0969; -.
DR TuberculList; Rv0943c; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 346 AA; 37205 MW; F51A5815B5A36136 CRC64;

Query Match          38.2%; Score 42; DB 1; Length 346;
Best Local Similarity 50.0%; Pred. No. 16;
Matches 8; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 2 PNVFIRAHDSVQTRIARI 17
      :||| | | : | | |

Query Match          37.3%; Score 41; DB 1; Length 160;
Best Local Similarity 42.1%; Pred. No. 10;
Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 NVVFIRAHDSVQTRIARI 21
Db 49 DYDAKAQOGGEIETRIARI 67
      :| :| | :||| :|

RESULT 14
GREA_MYCGE
ID GREA_MYCGE STANDARD; PRT; 161 AA.
AC P47524;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)

Db 177 PNFFITGPDTAQAAR 192
RESULT 13
GREA_MYCPN
ID GREA_MYCPN STANDARD; PRT; 160 AA.
AC P78019;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRANSCRIPTION ELONGATION FACTOR GREA (TRANSCRIPT CLEAVAGE FACTOR
DE GREA).
GN GREA OR MPN401 OR MP437.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA Herrmann R.;
RA "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- FUNCTION: NECESSARY FOR EFFICIENT RNA POLYMERASE TRANSCRIPTION
CC ELONGATION PAST TEMPLATE-ENCODED ARRESTING SITES. THE ARRESTING
CC SITES IN DNA HAVE THE PROPERTY OF TRAPPING A CERTAIN FRACTION
CC OF ELONGATING RNA POLYMERASES THAT PASS THROUGH, RESULTING IN
CC LOCKED TERNARY COMPLEXES. CLEAVAGE OF THE NASCENT TRANSCRIPT BY
CC CLEAVAGE FACTORS SUCH AS GREA OR GREB ALLOWS THE RESUMPTION OF
CC ELONGATION FROM THE NEW 3'TERMINUS. GREA RELEASES SEQUENCES OF
CC 2 TO 3 NUCLEOTIDES (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE GREA/GREB FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AE000043; AAB96085.1; -.
DR HSPB; P21346; IGRJ.
DR InterPro; IPR001437; Grea_Greb.
DR Pfam; PF01272; Grea_Greb; 1.
DR ProDom; PD004918; Grea_Greb; 1.
DR PROSITE; PS00829; GREAB.1; 1.
DR PROSITE; PS00830; GREAB.2; 1.
DR Transcription regulation; DNA-binding; Coiled coil; Complete proteome.
FT DOMAIN 8 28 COILED COIL (POTENTIAL).
FT DOMAIN 48 75 COILED COIL (POTENTIAL).
SQ SEQUENCE 160 AA; 18101 MW; 32448C91712A19C2 CRC64;

```

TRANSCRIPTION ELONGATION FACTOR GREB (TRANSCRIPT CLEAVAGE FACTOR  
DE GREA).  
GN GREA OR MG282.  
OS Mycoplasma genitalium.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 33530 / G-37;  
RA MEDLINE=96026346; PubMed=7569993;  
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,  
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,  
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,  
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,  
RA Tomb J.-F., Dougherty B.A., Brott K.F., Hu P.-C., Lucier T.S.,  
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;  
RA "The minimal gene complement of Mycoplasma genitalium.";  
RT Science 270:397-403(1995).  
RL CC  
CC -!- FUNCTION: NECESSARY FOR EFFICIENT RNA POLYMERASE TRANSCRIPTION  
CC ELONGATION PAST TEMPLATE-ENCODED ARRESTING SITES. THE ARRESTING  
CC SITES IN DNA HAVE THE PROPERTY OF TRAPPING A CERTAIN FRACTION  
CC OF ELONGATING RNA POLYMERASES THAT PASS THROUGH, RESULTING IN  
CC LOCKED TERNARY COMPLEXES. CLEAVAGE OF THE NASCENT TRANSCRIPT BY  
CC CLEAVAGE FACTORS SUCH AS GREA OR GREB ALLOWS THE RESUMPTION OF  
CC ELONGATION FROM THE NEW 3'TERMINUS. GREA RELEASES SEQUENCES OF  
CC 2 TO 3 NUCLEOTIDES (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE GREA/GREB FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U39708; AAC71504.1; -  
CC HSSP; P21346; 1GRJ.  
CC TIGR; MG282; -  
CC InterPro: IPR001437; Grea\_Greb.  
CC Pfam; PF01272; Grea\_Greb; 1.  
CC PRODOM; PD004918; Grea\_Greb; 1.  
CC PROSITE; PS00829; GREAB\_1; 1.  
CC PROSITE; PS00830; GREAB\_2; 1.  
CC Transcription regulation; DNA-binding; Coiled coil; Complete proteome.  
FT DOMAIN 9 28 COILED COIL (POTENTIAL).  
FT DOMAIN 47 76 COILED COIL (POTENTIAL).  
SQ SEQUENCE 161 AA; 18162 MW; 67E2850CB59ECC5B CRC64;  
Query Match 37.3%; Score 41; DB 1; Length 161;  
Best Local Similarity 42.1%; Pred. No. 10;  
Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;  
Qy 3 NYVFIRAHSDSEVQTRIAKI 21  
Db 49 DYDAKAQGGIEIETRIAI 67  
RESULT 15  
KAD4\_MOUSE  
ID KAD4\_MOUSE STANDARD; PRT; 223 AA.  
AC Q9WUR9; O9R1X7;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE ADENYLATE KINASE ISOENZYME 4, MITOCHONDRIAL (EC 2.7.4.3) (ATP-AMP  
DE TRANSPHOSPHORYLASE).  
GN AK4 OR AK-4 OR AK3B.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI\_TaxID=10090;  
OX [1]  
RN SEQUENCE FROM N.A.  
RP  
RC TISSUE=Brain;  
RA MEDLINE=99033072; PubMed=9813319;  
RA Yoneda T., Sato M., Maeda M., Takagi H.;  
RA "Identification of a novel adenylate kinase system in the brain:  
RT cloning of the fourth adenylate kinase.";  
RL Brain Res. Mol. Brain Res. 62:187-195(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Noma T.;  
RA Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
RL -!- CATALYTIC ACTIVITY: GTP + AMP = GDP + ADP.  
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX (POTENTIAL).  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN THE PYRAMIDAL CELLS IN THE  
CC HIPPOCAMPUS.  
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE CENTRAL NERVOUS SYSTEM IN A  
CC REGION-SPECIFIC MANNER FROM THE MIDDLE STAGE OF EMBRYOGENESIS TO  
CC THE ADULTHOOD IN THE RODENT.  
CC -!- SIMILARITY: BELONGS TO THE ADENYLATE KINASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; D85036; BAA77760.1; -  
CC EMBL; AB020239; BAA77363.1; -  
CC MGD; MGI-87980; AK4.  
CC InterPro: IPR000850; Adenylate\_kin.  
CC Pfam; PF00406; adenylatekinase; 1.  
CC PRINTS; PR00094; ADENYLTKINASE.  
CC PRODOM; PD000657; Adenylate\_kin; 1.  
CC PROSITE; PS00113; ADENYLATE\_KINASE; 1.  
CC Transferrase; Kinase; GTP-binding; Mitochondrion.  
FT NP\_BIND 12 20 GTP (BY SIMILARITY).  
FT CONFLICT 68 68 V -> A (IN REF. 2).  
FT CONFLICT 187 187 S -> N (IN REF. 2).  
FT CONFLICT 190 190 V -> M (IN REF. 2).  
SQ SEQUENCE 223 AA; 25061 MW; 50552294971515EC CRC64;  
Query Match 37.3%; Score 41; DB 1; Length 223;  
Best Local Similarity 45.0%; Pred. No. 15;  
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;  
Qy 1 VPNVVFIRAHSDSEVQTRIAK 20  
Db 64 VPDHVITRIMMSELETRSAQ 83  
Search completed: March 27, 2002, 14:27:08  
Job time: 1650 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:26:09 ; Search time 188.53 Seconds  
(without alignments)  
17.069 Million cell updates/sec

Title: US-09-290-049a-18

Perfect score: 110

Sequence: 1 VPNVVFIHSDSEVQTRIAKII 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	110	100.0	1338	2 Q9WXJ4	Q9WXJ4 streptococc
2	90	81.8	1477	2 Q9L466	Q9L466 leuconostoc
3	90	81.8	1508	2 O52224	O52224 leuconostoc
4	90	81.8	1508	2 Q9EZH5	Q9EZH5 leuconostoc
5	89	80.9	1527	2 Q9ZAR4	Q9ZAR4 leuconostoc
6	88	80.0	1512	2 Q9WXJ5	Q9WXJ5 streptococc
7	87	79.1	1016	2 Q9LCJ7	Q9LCJ7 leuconostoc
8	86	78.2	1290	2 Q48756	Q48756 leuconostoc
9	83	75.5	1575	2 Q9LCH3	Q9LCH3 streptococc
10	83	75.5	1577	2 Q54178	Q54178 streptococc
11	77	70.0	1577	2 Q55265	Q55265 streptococc
12	76	69.1	1455	2 Q69388	Q69388 streptococc
13	75	68.2	1390	2 Q69385	Q69385 streptococc
14	75	68.2	1455	2 Q69382	Q69382 streptococc
15	75	68.2	1455	2 Q69391	Q69391 streptococc
16	75	68.2	1455	2 Q69397	Q69397 streptococc
17	75	68.2	1599	2 Q00599	Q00599 streptococc
18	74	67.3	1449	2 Q68542	Q68542 streptococc
19	74	67.3	1449	2 Q55264	Q55264 streptococc

20	73	66.4	1518	2	Q00600	Q00600 streptococc
21	69	62.7	1590	2	Q55263	Q55263 streptococc
22	69	62.7	1590	2	Q59983	Q59983 streptococc
23	63	57.3	2057	2	Q9RE05	Q9RE05 leuconostoc
24	50	45.5	1442	10	Q9XHV5	Q9XHV5 oryza sativ
25	48	43.6	492	2	Q9WXX1	Q9WXX1 thermotoga
26	45.5	41.4	866	5	Q9VBT7	Q9VBT7 drosophila
27	45	40.9	190	2	Q85349	Q85349 caulobacter
28	45	40.9	498	2	Q9A9H5	Q9A9H5 caulobacter
29	44	40.0	583	10	Q49505	Q49505 arabidopsis
30	43.5	39.5	727	10	Q9SXG9	Q9SXG9 physcomitre
31	43	39.1	148	5	Q9NKU3	Q9NKU3 leishmania
32	43	39.1	210	5	Q9N6L5	Q9N6L5 leishmania
33	43	39.1	457	2	Q84126	Q84126 chlamydia t
34	43	39.1	457	2	Q9PKR4	Q9PKR4 chlamydia m
35	43	39.1	765	2	P74599	P74599 synechocyst
36	43	39.1	1070	2	Q55365	Q55365 synechocyst
37	43	39.1	2879	5	Q9U0Z1	Q9U0Z1 leishmania
38	42.5	38.6	319	2	Q34131	Q34131 lactococcus
39	42.5	38.6	319	2	Q9CG88	Q9CG88 lactococcus
40	42	38.2	144	2	Q9KAL5	Q9KAL5 bacillus ha
41	42	38.2	207	2	Q9L650	Q9L650 lactococcus
42	42	38.2	472	5	Q9WSW9	Q9WSW9 drosophila
43	42	38.2	481	2	Q9FAX0	Q9FAX0 flexibacter
44	42	38.2	481	2	Q9AQL4	Q9AQL4 chitinophag
45	42	38.2	922	3	Q9HEJ8	Q9HEJ8 neurospora

ALIGNMENTS

RESULT 1						
Q9WXJ4						
ID Q9WXJ4	PRELIMINARY;					
AC Q9WXJ4						PRT; 1338 AA.
DT 01-NOV-1999	(TRENBLrel. 12, Created)					
DT 01-NOV-1999	(TRENBLrel. 12, Last sequence update)					
DE 01-JUN-2001	(TRENBLrel. 17, Last annotation update)					
DE GTF-S.						
GN GTF-S.						
OS Streptococcus criceti.						
OC Plasmid pAM1.						
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;						
OC Streptococcus.						
OX NCBI_TaxID=1333;						
RN [1]						
RP SEQUENCE FROM N.A.						
RC STRAIN=HS-6;						
RA Inoue M., Fukui K., Miyagi A.;						
RT "S.cricetus glucosyltransferase(gtfs and gtf) genes."						
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.						
DR EMBL; AB026123; BAA77236.1; -						
DR HSSP; P06278; IVJS.						
DR InterPro; IPR002479; CW_binding.						
DR InterPro; IPR003318; Glyco_hydro_70.						
DR Pfam; PF01473; CW_binding_1; 10.						
DR Pfam; PF02324; Glyco_hydro_70; 1.						
KW Plasmid.						
Q9WXJ4	SEQUENCE 1338 AA; 148558 MW; 0A90C8E10E15D99B CRC64;					
Query Match	100.0%;	Score 110;	DB 2;	Length 1338;		
Best Local Similarity	100.0%;	Pred. No. 1.7e-09;				
Matches 22;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		
Qy 1 VPNVVFIHSDSEVQTRIAKII 22						
Db 509 VPNVVFIHSDSEVQTRIAKII 530						
RESULT 2						
Q9L466						
ID Q9L466	PRELIMINARY;					PRT; 1477 AA.

Q9L466;  
AC 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE DEXTRANSUCRASE (EC 2.4.1.5).  
GN DSR.  
OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OC Leuconostoc.  
OX NCBI\_TaxID=1245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NRRL B-1355;  
RA Arquello-Morales M.A., Renaud-Simeon M., Pizzut S., Sarcabal P.,  
RW Willemot R.M., Monsan P.;  
RT "Sequence analysis of the gene encoding alternansucrase, a sucrose  
glucosyltransferase from Leuconostoc mesenteroides NRRL B-1355.";  
RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
EMBL: AJ250172; CAB76565.1; -;  
InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam: PF01473; CW\_binding\_1; 14.  
DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
KW Transferase; Glycosyltransferase.  
SQ SEQUENCE 1477 AA; 164886 MW; E6F5710DEDFCB831 CRC64;

Query Match 81.8%; Score 90; DB 2; Length 1477;  
Best Local Similarity 77.3%; Pred. No. 3.9e-06;  
Matches 17; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPNVVFIKRAHDSVQVTRIAKII 22  
:|||||:|||||:|||||:|||||:|||||  
DB 603 IPNYSFVRAHDSVQVTRIAQII 624

RESULT 3  
O52224  
ID O52224 PRELIMINARY; PRT; 1508 AA.  
AC O52224;  
DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE 6-  
GLUCOSYLTRANSFERASE).  
GN DSRB.  
OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OC Leuconostoc.  
OX NCBI\_TaxID=1245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NRRL B-1299;  
RA Monchois V., Renaud-Simeon M., Monsan P., Willemot R.M.;  
RW FEMS Microbiol. Lett. 0:0-0(1998).  
RL C-1- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) = D-  
FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).  
CC EMBL: AF030129; AAB95453.1; -;  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam: PF01473; CW\_binding\_1; 14.  
DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
KW Transferase; Glycosyltransferase.  
SQ SEQUENCE 1508 AA; 168511 MW; E70CEB57A70D1F0 CRC64;

Query Match 81.8%; Score 90; DB 2; Length 1508;  
Best Local Similarity 77.3%; Pred. No. 4e-06;  
Matches 17; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPNVVFIKRAHDSVQVTRIAKII 22  
:|||||:|||||:|||||:|||||:|||||  
DB 634 IPNYSFVRAHDSVQVTRIAQII 655

RESULT 4  
Q9EZHS  
ID Q9EZHS PRELIMINARY; PRT; 1508 AA.  
AC Q9EZHS;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE DEXTRANSUCRASE DSRB742.  
GN DSRB742.  
OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OC Leuconostoc.  
OX NCBI\_TaxID=1245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B-742CB;  
RA Kim H.-S., Kim D., Ryu H.-J., Robyt J.F.;  
RW "Leuconostoc mesenteroides B-742CB, a dextranucrase gene.";  
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
EMBL: AF294469; AAG38021.1; -;  
DR EMBL: AF294469; AAG38021.1; -;  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam: PF01473; CW\_binding\_1; 14.  
DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
SQ SEQUENCE 1508 AA; 168542 MW; E2FCFA0F87AE4F3A CRC64;

Query Match 81.8%; Score 90; DB 2; Length 1508;  
Best Local Similarity 77.3%; Pred. No. 4e-06;  
Matches 17; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPNVVFIKRAHDSVQVTRIAKII 22  
:|||||:|||||:|||||:|||||:|||||  
DB 634 IPNYSFVRAHDSVQVTRIAQII 655

RESULT 5  
Q9ZAR4  
ID Q9ZAR4 PRELIMINARY; PRT; 1527 AA.  
AC Q9ZAR4;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE DEXTRANSUCRASE.  
GN DEX.  
OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OC Leuconostoc.  
OX NCBI\_TaxID=1245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NRRL B-512-F;  
RA Bhatnagar R., Singh D.K.S.;  
RW "Cloning and Molecular Characterization of Dextranucrase Gene from  
Leuconostoc mesenteroides NRRL B-512F.";  
RT Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
EMBL: U81374; AAD10952.1; -;  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam: PF01473; CW\_binding\_1; 16.  
DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
SQ SEQUENCE 1527 AA; 169709 MW; 1DFAFA237C743398 CRC64;

Query Match 80.9%; Score 89; DB 2; Length 1527;  
Best Local Similarity 72.7%; Pred. No. 6e-06;  
Matches 16; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPNVVFIKRAHDSVQVTRIAKII 22  
:|||||:|||||:|||||:|||||:|||||  
DB 652 IPNYSFVRAHDSVQVTRIAQIV 673



```
RESULT 6
Q9WXJ5 PRELIMINARY; PRT; 1512 AA.
AC Q9WXJ5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GTF-S.
GN GTF.
OS Streptococcus criceti.
OG Plasmid pAMI.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1333;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HS-6;
RA Inoue M., Fukui K., Miyagi A.;
RT "S. cricetus glucosyltransferase(gtfs and gtfT) genes.";
LN Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
EMBL: AB026123; BAA77237.1;
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1; 14.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Plasmid.
SQ SEQUENCE 1512 AA; 167145 MW; 4C03D9CB501FC14 CRC64;

Query Match 80.0%; Score 88; DB 2; Length 1512;
Best Local Similarity 81.0%; Pred. No. 8.6e-06;
Matches 17; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 2 PNYVFIHSDSEVQTRIAKII 22
I:|||||
Db 560 PSYVFVRAHSDSEVQTVIAQII 580

RESULT 7
Q9LCJ7 PRELIMINARY; PRT; 1016 AA.
AC Q9LCJ7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DEXTRANSUCRASE.
GN DSRT.
OS Leuconostoc mesenteroides.
OG Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL B-512F;
RA Funane K., Mizuno K., Takahara H., Kobayashi M.;
RT "Gene encoding a dextranucrase-like protein in Leuconostoc mesenteroides NRRL B-512F.";
LN Biosci. Biotechnol. Biochem. 64:29-38(2000).
EMBL: AB020020; BAA90527.1;
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1016 AA; 110343 MW; 8896EFD13CCCB47 CRC64;

Query Match 79.1%; Score 87; DB 2; Length 1016;
Best Local Similarity 81.0%; Pred. No. 8.2e-06;
Matches 17; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 PNYVFIHSDSEVQTRIAKII 22
I:|||||
```

```
Db 625 PNYSFVRAHSDSEVQTVIAEII 645

RESULT 8
Q48756 PRELIMINARY; PRT; 1290 AA.
AC Q48756;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DEXTRANSUCRASE.
OS Leuconostoc mesenteroides.
OG Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL B1299;
RA Monchois V., Willemot R.M., Renaud-Simeon M., Croux C., Monsan P.;
RT "Cloning and sequencing of a gene coding for a novel dextranucrase from Leuconostoc mesenteroides NRRL B-1299 synthesizing only alpha (1-6) and alpha (1-3) linkages.";
LN Gene 182:23-32(1996).
EMBL: U38181; AAB40875.1;
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1; 11.
DR Pfam: PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1290 AA; 145590 MW; 3555C2E96B749FAA CRC64;

Query Match 78.2%; Score 86; DB 2; Length 1290;
Best Local Similarity 85.7%; Pred. No. 1.6e-05;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 PNYVFIHSDSEVQTRIAKII 22
I:|||||
Db 388 PNYSFVRAHSDSEVQTVIAEII 408

RESULT 9
Q9LCH3 PRELIMINARY; PRT; 1575 AA.
AC Q9LCH3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE.
GN GTPR.
OS Streptococcus oralis.
OG Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC10557;
RA Fujiwara T., Hoshino T., Ooshima T., Sobue S., Hamada S.;
RT "Purification, characterization, and molecular analysis of the gene encoding glucosyltransferase from Streptococcus oralis.";
LN Infect. Immun. 68:2475-2483(2000).
EMBL: AB025228; BAA95201.1;
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1; 17.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1575 AA; 176792 MW; 772A26EAD7C2E543 CRC64;

Query Match 75.5%; Score 83; DB 2; Length 1575;
Best Local Similarity 80.0%; Pred. No. 6.1e-05;
```

```
Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 3 NVVFIKRAHDSVQVTRIAKII 22
DB 619 NVVFIKRAHDSVQVTRIAKII 638

RESULT 10
Q54178 Q54178 PRELIMINARY; PRT: 1577 AA.
AC Q54178; Q54247;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE.
GN GTF6.
OS Streptococcus gordonii challis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
NCBI_TaxID=29390;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CHALLIS;
RX MEDLINE=96157084; PubMed=8586195;
RA Vickerman M.M., Sulavik M.C., Clewell D.B.;
RT "Molecular analysis of Streptococcus gordonii glucosyltransferase
phase variants."
RL Dev. Biol. Stand. 85:309-314(1995).
RN [2]
RP SEQUENCE OF 1-96 FROM N.A.
RC STRAIN=CHALLIS;
RX MEDLINE=92276337; PubMed=1534326;
RA Sulavik M.C., Tardif G., Clewell D.B.;
RT "Identification of a gene, rgs, which regulates expression of
glucosyltransferase and influences the Spp phenotype of Streptococcus
gordonii Challis."
RL J. Bacteriol. 174:3577-3586(1992).
RN [3]
DR EMBL; U12643; AAC43483.1; -
DR EMBL; M89776; AAA26969.1; -
DR InterPro; IPR002479; CW_binding.
DR Pfam; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 18.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1577 AA; 177805 MW; 5AE0328DC5E08D18 CRC64;

Query Match 75.5%; Score 83; DB 2; Length 1577;
Best Local Similarity 80.0%; Pred. No. 6.1e-05;
Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 3 NVVFIKRAHDSVQVTRIAKII 22
DB 621 NVVFIKRAHDSVQVTRIAKII 640

RESULT 11
Q55265 Q55265 PRELIMINARY; PRT: 1577 AA.
AC Q55265;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE PRECURSOR.
GN GTFW.
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95122197; PubMed=7822030;
RA Simpson C.L., Giffard P.M., Jacques N.A.;
```

```
RT Streptococcus salivarius ATCC 25975 possesses at least two genes
RT coding for primer-independent glucosyltransferases."
RL Infect. Immun. 63:609-621(1995).
DR EMBL; L35928; AAC41413.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 11.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Signal; Transferase.
FT SIGNAL 1 38 POTENTIAL.
FT CHAIN 39 1577 GLUCOSYLTRANSFERASE.
SQ SEQUENCE 1577 AA; 175290 MW; 3EFB898A7D3A7BF3 CRC64;

Query Match 70.0%; Score 77; DB 2; Length 1577;
Best Local Similarity 70.0%; Pred. No. 0.00061;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 3 NVVFIKRAHDSVQVTRIAKII 22
DB 661 NVVFIKRAHDSVQVTRIAKII 680

RESULT 12
O69388 O69388 PRELIMINARY; PRT: 1455 AA.
AC O69388;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTF6.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WT4239;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
Streptococcus mutans."
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D88658; BAA26110.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 10.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1455 AA; 163045 MW; 6D90A4978D35DD82 CRC64;

Query Match 69.1%; Score 76; DB 2; Length 1455;
Best Local Similarity 72.7%; Pred. No. 0.00082;
Matches 16; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 1 VFNVFIKRAHDSVQVTRIAKII 22
DB 578 VFNVFIKRAHDSVQVTRIAKII 599

RESULT 13
O69385 O69385 PRELIMINARY; PRT: 1390 AA.
AC O69385;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTF6.
OS Streptococcus mutans.
```

```
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OX Streptococcus.
RN NCBI_TaxID=1309;
RP SEQUENCE FROM N.A.
RC STRAIN=MT4245;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D88655; BAA26106.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding.1; 7.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1390 AA; 155375 MW; 8847E4956EF05E9F CRC64;

Query Match 68.2%; Score 75; DB 2; Length 1390;
Best Local Similarity 72.7%; Pred. No. 0.0011;
Matches 16; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 VPNVVFIRAHDSVQTRIAKII 22
Db 578 VPSYSFIRAHDSVQDLIRDI 599

RESULT 14
O69382
ID O69382 PRELIMINARY; PRT; 1455 AA.
AC O69382;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTFC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RP SEQUENCE FROM N.A.
RC STRAIN=MT8148;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D88652; BAA26102.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding.1; 9.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1455 AA; 162969 MW; 27D4D3A1EECA2939 CRC64;

Query Match 68.2%; Score 75; DB 2; Length 1455;
Best Local Similarity 72.7%; Pred. No. 0.0012;
Matches 16; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 VPNVVFIRAHDSVQTRIAKII 22
Db 578 VPSYSFIRAHDSVQDLIRDI 599

RESULT 15
O69391
ID O69391 PRELIMINARY; PRT; 1455 AA.
```

```
AC O69391;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTFC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT4251;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D88661; BAA26114.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding.1; 9.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1455 AA; 162804 MW; 683A359D873E9E1A CRC64;

Query Match 68.2%; Score 75; DB 2; Length 1455;
Best Local Similarity 72.7%; Pred. No. 0.0012;
Matches 16; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 VPNVVFIRAHDSVQTRIAKII 22
Db 578 VPSYSFIRAHDSVQDLIRDI 599

Search completed: March 27, 2002, 14:26:10
Job time: 1682 sec
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:58:01 ; Search time 198.55 Seconds  
(without alignments)  
8.208 Million cell updates/sec

Title: US-09-290-049a-19

Perfect score: 110

Sequence: 1 VPSYFARAHDSVQDIIRDII 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_1101.\*  
1: /SIDSB/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDSB/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDSB/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDSB/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDSB/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDSB/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDSB/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDSB/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDSB/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDSB/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDSB/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDSB/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDSB/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDSB/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDSB/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDSB/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDSB/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDSB/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDSB/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDSB/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDSB/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDSB/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	100.0	1592	14	AA19925
2	68	61.8	2057	21	AA10667
3	64	58.2	1577	17	AA191047
4	46	41.8	519	20	AA148586
5	46	41.8	743	22	AA193069
6	46	41.8	743	22	AA193298
7	45.5	41.4	484	18	AA121897
8	45	40.9	541	8	AA170433
9	45	40.9	541	17	AA195586
10	44.5	40.5	451	22	AA141255
11	43	39.1	2625	19	AA155887

12	43	39.1	2627	19	AAW61347	Human telomerase R
13	42	38.2	85	21	AAG28140	Arabidopsis thalia
14	42	38.2	91	21	AAG28139	Arabidopsis thalia
15	42	38.2	392	20	AAW97743	Soybean S-adenosyl
16	42	38.2	480	14	AAAR2867	Platelet aggregati
17	42	38.2	537	8	AAAP70432	Bovine atrial natri
18	42	38.2	537	17	AAAR9585	Atrial natriuretic
19	41	37.3	164	21	AAAG44998	zea mays protein f
20	41	37.3	166	22	AAAK36026	Peptide #10063 enc
21	41	37.3	190	21	AAAG44997	zea mays protein f
22	41	37.3	210	21	AAAG44996	zea mays protein f
23	41	37.3	217	22	AAU04854	Micromonospora eve
24	41	37.3	267	22	AAU00297	Interferon induced
25	41	37.3	271	22	AAU00299	Interferon induced
26	41	37.3	304	22	AAU00296	Interferon induced
27	41	37.3	308	22	AAU00298	Interferon induced
28	41	37.3	366	21	AAAG35916	zea mays protein f
29	41	37.3	639	18	AAW26530	Trypanosoma cruzi
30	41	37.3	639	20	AAAT23292	Trypanosoma cruzi
31	41	37.3	1684	12	AAAR14948	Bacterial amylase
32	40.5	36.8	92	22	AAG76844	Human colon cancer
33	40.5	36.8	348	18	AAW27267	Streptococcus pneu
34	40.5	36.8	348	20	AAAT29981	Streptococcus pneu
35	40.5	36.8	480	20	AAU09552	Streptococcus pneu
36	40.5	36.8	480	21	AAAT28229	Glutamyl trNA synt
37	40.5	36.8	573	21	AAAG29545	Arabidopsis thalia
38	40.5	36.8	576	21	AAAG29544	Arabidopsis thalia
39	40.5	36.8	604	21	AAAG29543	Arabidopsis thalia
40	40	36.4	126	21	AAAB26249	Corn cyclin-depend
41	40	36.4	164	20	AAAT45220	Human CASB47 deriv
42	40	36.4	221	20	AAAT45219	Human CASB47 prote
43	40	36.4	252	19	AAAB38529	Streptococcus pneu
44	40	36.4	323	18	AAAW55360	H. pylori ORF 03ge
45	40	36.4	323	20	AAAT71165	H. pylori outer me

#### ALIGNMENTS

RESULT 1  
AA132925  
ID AA132925 standard; Protein; 1592 AA.  
XX  
AC AA132925;  
XX  
DT 28-JUN-1993 (first entry)  
XX  
DE Glucosyltransferase I.  
XX  
KW GT-1; Streptococcus; dental; caries.  
XX  
OS Streptococcus sobrinus.  
XX  
FN JP05023188-A.  
XX  
PD 02-FEB-1993.  
XX  
PF 25-JUL-1991; 91JP-0186592.  
XX  
PR 25-JUL-1991; 91JP-0186592.  
XX  
PA (FUKU/) FUKUI I.  
PA (KATO/) KATO K.  
XX  
DR WPI; 1993-079449/10.  
XX  
PT N-PSDB; AAQ37760.  
XX  
PT DNA sequence glucosyl:transferase-I - comprises Streptococcus  
PT sobrinus DNA sequence with at least one nucleotide added or  
PT deleted  
XX  
PS Claim 13; Page 15; 29pp; Japanese.  
XX

CC The DNA sequence from Streptococcus sobrinus strain 6715 encodes  
 CC glucosyltransferase-I (and mutants). The DNA was obtd. by treating  
 CC S. sobrinus 6715 with mutanolysin, extracting the chromosomal DNA,  
 CC partially digesting with Sau3AI and fractionating on agarose gel.  
 CC The 3-5 kbp fragment was ligated into pUC18 and E. coli JM109  
 CC transformed with it. A Gt-I expressing clone was isolated and  
 CC sequenced. The clone may be used in the development of a drug for  
 CC dental caries.  
 XX SQ Sequence 1592 AA;

Query Match 100.0%; Score 110; DB 14; Length 1592;  
 Best Local Similarity 100.0%; Pred. No. 6e-10;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QV 1 VPSYSFARAHSEVODIIRII 22  
 |||||  
 548 vpsysfarahdsevdqirdii 569

RESULT 2

AAB10667  
 ID AAB10667 standard; Protein; 2057 AA.  
 AC AAB10667;  
 XX 19-JAN-2001 (first entry)  
 DT  
 XX L. mesenteroides alternan sucrose protein.  
 DE  
 DE Alternan sucrose; glucosyltransferase; fructose; cosmetic; foodstuff;  
 KW syrup.  
 KW  
 KW  
 XX Leuconostoc mesenteroides.  
 OS  
 XX DE19905069-A1.  
 PN  
 XX 10-AUG-2000.  
 PD  
 XX 08-FEB-1999; 99DE-1005069.  
 PF  
 XX 08-FEB-1999; 99DE-1005069.  
 PR  
 XX (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.  
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
 PI  
 XX Kossmann J, Welsh T, Quanz M, Knuth K;  
 XX WPI; 2000-550294/51.  
 DR N-PSDB; AAA97904.  
 DR  
 XX New nucleic acid encoding recombinant Leuconostoc mesenteroides  
 PT alternan sucrose protein and methods of alternan and fructose  
 PT production  
 PT  
 XX Claim 1a; Page 30-36; 64pp; German.  
 PS  
 XX This invention describes a novel nucleic acid molecule (I) encoding an  
 CC alternan sucrose (E.C. 2.4.1.140 - an enzyme, that belongs to the  
 CC glucosyltransferase group) The recombinant, purified alternan sucrose  
 CC gene is useful for the fermentative production of alternan (a  
 CC carbohydrate) and/or fructose by secreting the enzyme into a  
 CC saccharose-containing culture medium. Alternatively, the enzyme is  
 CC contacted with a saccharose-containing solution. The alternan and/or  
 CC fructose is then isolated from the medium. Cosmetic products or  
 CC foodstuffs containing alternan can be produced. Recombinant production of  
 CC alternan sucrose is advantageous as it provides a cost effective means of  
 CC producing fructose for high fructose containing syrups, production of  
 CC starch. This sequence represents the Leuconostoc mesenteroides alternan  
 CC sucrose protein which is described in the method of the invention.  
 XX

SQ Sequence 2057 AA;

Query Match 61.8%; Score 68; DB 21; Length 2057;  
 Best Local Similarity 63.2%; Pred. No. 0.013;  
 Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QV 1 VPSYSFARAHSEVODIIR 19  
 :||:|||||:|||||  
 Db 757 ipnysfvrahdydadpdr 775

RESULT 3

AAR91047  
 ID AAR91047 standard; Protein; 1577 AA.  
 XX  
 AC AAR91047;  
 XX 22-MAY-1996 (first entry)  
 DT  
 XX Alpha-D-glucosyltransferase.  
 DE  
 XX Alpha-D-glucosyltransferase; primer-independent; soluble glucan;  
 KW sucrose; transgenic plant; cloning; Escherichia coli;  
 KW phage lambda-C13; vector; plasmid pGS501; plasmid pGS502;  
 KW gene transfer; crop improvement; storage carbohydrate; pasture;  
 KW feedstuff; senescence; dextran; binder; food; pharmaceutical.  
 XX  
 OS Streptococcus salivarius strain ATCC 25975.  
 XX  
 XX WO9606173-A1.  
 PN  
 XX 29-FEB-1996.  
 PD  
 XX 24-AUG-1995; 95WO-AU00527.  
 PF  
 XX 24-AUG-1994; 94AU-0007643.  
 PR  
 XX (GIFF/) GIFFARD P M.  
 PA (JACQ/) JACQUES N A.  
 PA (SIMP/) SIMPSON C L.  
 XX  
 XX Giffard PM, Jacques NA, Simpson CL;  
 PI  
 XX WPI; 1996-151376/15.  
 DR N-PSDB; AAT13139.  
 DR  
 XX Plants contg. new bacterial DNA encoding glucosyl transferase  
 PT activity - retain higher levels of stored carbohydrate(s) in a form  
 PT readily digestible by ruminants  
 PT  
 XX Claim 4; Page 16-20; 31pp; English.  
 PS  
 XX The sequence represents an alpha-D-glucosyltransferase from  
 CC Streptococcus salivarius. The enzyme is primer-independent, and  
 CC produces soluble glucan from sucrose. A gene encoding the enzyme  
 CC may be cloned and expressed in Escherichia coli using a subclone  
 CC of phage lambda-C13, e.g. plasmid pGS501 or plasmid pGS502. The  
 CC DNA may also be expressed in a transgenic plant, to improve the  
 CC level of stored carbohydrate in a pasture plant, which normally  
 CC contains low levels, or to prevent degradation of stored carbohydrate  
 CC during plant senescence. Dextran may be isolated from the plant, for  
 CC use as a food binder or pharmaceutical additive. Primer independence  
 CC ensures that the enzyme will be functional in plants. The glucan is  
 CC poorly degraded in plants but easily degraded by bacteria in the rumen  
 CC of grazing livestock.  
 XX  
 XX SQ Sequence 1577 AA;

Query Match 58.2%; Score 64; DB 17; Length 1577;  
 Best Local Similarity 60.0%; Pred. No. 0.045;  
 Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 3 SYSFARAHSEVQDIIRDII 22  
: | | | | | | | : : | |  
Db 661 nyifvrahdsqvavlanii 680

RESULT 4  
AAy48586  
ID AAY48586 standard; Protein; 519 AA.

XX AAY48586;  
XX 08-DEC-1999 (first entry)  
XX Human breast tumour-associated protein 47.

XX Expressed sequence tag; EST; human; breast; cancer; gene therapy;  
KW treatment; tumour; cytostatic; medicament.  
XX Homo sapiens.

OS DE19813839-A1.  
XX 23-SEP-1999.

XX 20-MAR-1998; 98DE-1013839.  
XX 20-MAR-1998; 98DE-1013839.

XX (META-) METAGEN GES GENOMFORSCHUNG MBH.  
XX Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E, Rosentahl A;

XX WPI; 1999-528981/45.  
XX N-PSDB; AA233656.

XX Human nucleic acid sequences and protein products from tumor breast  
PT tissue, useful for breast cancer therapy -  
XX Claim 22; 163; 188pp; German.

XX This invention describes novel human nucleic acid sequences from tumor  
CC breast tissue which have cytostatic activity. The nucleic acid sequences  
CC can be used to produce and isolate full-length gene sequences. They can  
CC be used to express proteins, which can be used as tools to find an  
CC activity against breast cancer. The sequences can be used in sense or  
CC antisense form. They are especially useful for medicaments for gene  
CC therapy to treat breast cancer. AAY48540-148617 represent protein  
CC fragments encoded by the expressed sequence tags described in the method  
CC of the invention.

Sequence 519 AA;

Query Match 41.8%; Score 46; DB 20; Length 519;  
Best Local Similarity 36.4%; Pred. No. 15;  
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHSEVQDIIRDII 22  
| | | | | | | : : | |  
Db 469 vpqyfsnlppskdilqgvv 490

RESULT 5  
AAB93069  
ID AAB93069 standard; Protein; 743 AA.

XX AAB93069;  
XX 26-JUN-2001 (first entry)  
XX Human protein sequence SEQ ID NO:11889.

XX DE  
XX

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX Homo sapiens.  
XX EPI074617-A2.  
XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.  
XX 29-JUL-1999; 99JP-0248036.  
XX 27-AUG-1999; 99JP-0300253.  
XX 11-JAN-2000; 2000JP-0118776.  
XX 02-MAY-2000; 2000JP-0183767.  
XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX Claim 8; SEQ ID 11889; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

XX Sequence 743 AA;

Query Match 41.8%; Score 46; DB 22; Length 743;  
Best Local Similarity 36.4%; Pred. No. 23;  
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHSEVQDIIRDII 22  
| | | | | | | : : | |  
Db 693 vpqyfsnlppskdilqgvv 714

RESULT 6  
AAB93298  
ID AAB93298 standard; Protein; 743 AA.

XX AAB93298;

XX 26-JUN-2001 (first entry)  
XX

DE Human protein sequence SEQ ID NO:12363.  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
KW Homo sapiens.  
XX EP1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-0116126.  
XX 29-JUL-1999; 99JP-0248036.  
XX 27-AUG-1999; 99JP-0300253.  
XX 11-YAN-2000; 2000JP-0118776.  
XX 02-MAY-2000; 2000JP-0183767.  
XX 09-JUN-2000; 2000JP-0241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
XX full-length cDNAs defined in the specification, and for the detection  
XX and/or diagnosis of the abnormality of the proteins encoded by the  
XX full-length cDNAs -  
XX Claim 8; SEQ ID 12363; 2537pp + CD ROM; English.  
XX The present invention describes primer sets for synthesizing 5602  
XX full-length cDNAs defined in the specification. Where a primer set  
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
XX to the complementary strand of a polynucleotide which comprises one of  
XX the 5602 nucleotide sequences defined in the specification, where the  
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
XX of an oligonucleotide comprising a sequence complementary to the  
XX complementary strand of a polynucleotide which comprises a 5'-end  
XX sequence and an oligonucleotide comprising a sequence complementary to a  
XX polynucleotide which comprises a 3'-end sequence, where the  
XX oligonucleotide comprises at least 15 nucleotides and the combination of  
XX the 5'-end sequence/3'-end sequence is selected from those defined in  
XX the specification. The primer sets can be used in antisense therapy and  
XX in gene therapy. The primers are useful for synthesizing polynucleotides,  
XX particularly full-length cDNAs. The primers are also useful for the  
XX detection and/or diagnosis of the abnormality of the proteins encoded by  
XX the full-length cDNAs. The primers allow obtaining of the full-length  
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
XX represent oligonucleotides, all of which are used in the exemplification  
XX of the present invention.  
XX Sequence 743 AA;  
XX  
XX Query Match 41.8%; Score 46; DB 22; Length 743;  
XX Best Local Similarity 36.4%; Pred. No. 23;  
XX Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;  
XX  
XX Qy 1 VPSYSFARAHDEVDIIRDII 22  
XX ||||| :||:||||:  
XX Db 693 vpsyyfslpsskdlqqvv 714  
XX  
XX RESULT 7  
XX AA21897  
XX ID AA21897 standard; Protein; 484 AA.  
XX XX  
XX AC AA21897;  
XX XX

DT 11-SEP-1997 (first entry)  
XX Glutamyl-tRNA synthetase from Staph. aureus.  
DE Glutamyl-tRNA synthetase from Staph. aureus.  
XX tRNA synthetase; immunological response; antibody; protection;  
KW bacterial infection; adherence; damaged tissue; wound healing;  
KW skin; vaccine.  
XX Staphylococcus aureus.  
OS Staphylococcus aureus.  
XX EP785261-A1.  
XX 23-JUL-1997.  
PD 17-JAN-1997; 97EP-0300310.  
XX 19-JAN-1996; 96GB-0001069.  
PF 19-JAN-1996; 96GB-0001069.  
XX (SMIK ) SMITHKLINE BEECHAM PLC.  
XX Hodgson JE, Lawlor EJ;  
PI WPI; 1997-365936/34.  
XX N-PSDB; AAT73695.  
DR DNA encoding glutamyl-tRNA synthetase from Staphylococcus aureus  
XX WCUH 29 - useful for protection against bacterial infections  
XX Claim 1; Page 19-21; 31pp; English.  
XX  
XX The present sequence represents a novel glutamyl-tRNA synthetase protein  
XX from Staphylococcus aureus strain WCUH29 (NCIMB 40771). Vectors  
XX comprising the DNA (or polynucleotides having at least 70 % identity  
XX to it) can be used for the recombinant production of the enzyme. The  
XX enzyme or its related DNA (through gene therapy) is used to induce an  
XX immunological response in a mammal to generate antibodies to protect  
XX against disease. The antibodies protect against invasion of bacteria,  
XX e.g. by blocking adherence of bacteria to damaged tissue, including  
XX wounds in skin or connective tissue caused by mechanical, chemical or  
XX thermal damage or by implantation of in-dwelling devices, or wounds  
XX in the mucous membranes. Antagonists are used to inhibit the enzyme,  
XX especially to prevent adhesion of bacteria to mammalian extracellular  
XX matrix proteins on in-dwelling devices or to extracellular matrix  
XX proteins in wounds, or to block tRNA synthetase protein mediated  
XX mammalian cell invasion by, e.g. initiating phosphorylation of mammalian  
XX tyrosine kinase. Analysing a sample for the presence of the enzyme (or  
XX a polypeptide having at least 70 % identity to it) is used for a  
XX diagnostic process.  
XX Sequence 484 AA;  
XX  
XX Query Match 41.4%; Score 45.5; DB 18; Length 484;  
XX Best Local Similarity 40.9%; Pred. No. 17;  
XX Matches 9; Conservative 6; Mismatches 4; Indels 3; Gaps 1;  
XX  
XX Qy 1 VPSYSFARA---HDSEVDIIR 19  
XX :||:||||| :||:|||||  
XX Db 193 iptynfavaldhymqisdvir 214  
XX  
XX RESULT 8  
XX AAP70433  
XX ID AAP70433 standard; protein; 541 AA.  
XX XX  
XX AC AAP70433;  
XX XX  
XX 27-JAN-1991 (first entry)  
XX Human atrial natriuretic factor receptor.  
DE Atrial natriuretic factor receptor; hypertension.  
XX Atrial natriuretic factor receptor; hypertension.  
XX XX



OS Homo sapiens.

XX WO8706938-A.

XX 19-NOV-1987.

XX PF 11-MAY-1987; 87WO-US01122.

XX PR 09-MAY-1986; 86US-0861529.

XX PA (BIOT-) BIOTECHNOL RES ASSO.

XX PI Schenk DB;

XX DR WPI; 1987-334947/47.

XX DR N-PSDB; AAN70692.

XX PT Purified atrial natriuretic receptor peptide - for diagnosing, hypertension, and corresp. DNA coding sequences, recombinant PT vectors, transformed cells and antibodies, etc.

XX PT Disclosure; fig 5; 62pp; English.

CC This atrial natriuretic factor (ANP) receptor is used to diagnose hypertension, test ANP analogues for binding/blocking activity and for the therapeutic reduction of serum ANP levels. Antibodies CC raised against this protein can be used for purifying the receptor CC and for blocking binding of the ANP protein to its receptor. CC See also AAN70691.

XX SQ Sequence 541 AA;

Query Match 40.9%; Score 45; DB 8; Length 541;  
Best Local Similarity 44.4%; Pred. No. 23;  
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 4 YSFARAHDSVQDIIRDI 21

DB 234 ysfdetkdldldilrni 251

RESULT 9

AAR95586

ID AAR95586 standard; Protein; 541 AA.

XX: AAR95586;

XX DT 11-OCT-1996 (first entry)

XX Atrial natriuretic peptide receptor.

KW Human; atrial natriuretic peptide receptor; placenta;  
KW signal peptide; hypertension; diagnostic; hypotensive; antibody;  
KW immunoassay; affinity chromatography; purification.

XX Homo sapiens.

XX US5512455-A.

XX PD 30-APR-1996.

XX PF 09-MAY-1986; 86US-0861529.

XX PR 11-MAY-1987; 87US-0048296.

XX PR 09-MAY-1986; 86US-0861529.

XX PA (SCIO-) SCIOS NOVA INC.

XX PI Schenk DB;

XX DR WPI; 1996-229865/23.

XX DR N-PSDB; AAT27029.

XX

PT New isolated atrial natriuretic peptide receptor DNA - used for the  
PT prodn. of ANP receptor proteins for use in e.g. diagnosis, therapy  
PT or antibody prodn.

XX PS Example 3; Fig 5; 36pp; English.

XX CC The sequence represents a human atrial natriuretic peptide (ANP)  
CC receptor, isolated from placenta tissue. A cDNA encoding the  
CC receptor has been cloned in Escherichia coli, and the

CC encoded protein shows homology to the bovine aorta smooth muscle  
CC ANP receptor (AAR95585). The receptor may be used to determine ANP  
CC levels, e.g. in hypertension diagnosis, or therapeutically to  
CC reduce ANP levels. Antibodies against the protein may be used as  
CC diagnostic agents or in purification of ANP receptor.

XX SQ Sequence 541 AA;

Query Match 40.9%; Score 45; DB 17; Length 541;  
Best Local Similarity 44.4%; Pred. No. 23;  
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 4 YSFARAHDSVQDIIRDI 21

DB 234 ysfdetkdldldilrni 251

RESULT 10

AAM41255

ID AAM41255 standard; Protein; 451 AA.

XX: AAM41255;

XX DT 22-OCT-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 6186.

XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.

XX OS Homo sapiens.

XX PN WO200153312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2001; :000WO-US34263.

XX PR 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-0552317.

XX PR 09-JUL-2000; 2000US-0598042.

XX PR 19-JUL-2001; 2000US-0620312.

XX PR 03-AUG-2000; 2000US-0653450.

XX PR 14-SEP-2000; 2000US-0662191.

XX PR 19-OCT-2000; 2000US-0693036.

XX PR 29-NOV-2000; 2000US-0727344.

XX PA (HYSE-) HYSEO INC.

XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX DR N-PSDB; AAI60411.

PT Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -



CC such as in cancer cells in which TRIP1 activity and/or TP2 activity  
CC is elevated, TRIP1 and/or TP2 may serve as a target to identify a  
CC molecule which inhibits activity, or which decreases or inhibits  
CC the protein-protein interaction of TRIP1 and TP2, or the binding of  
CC either TRIP1 or TP2 to telomerase RNA. Alternatively, ex vivo or  
CC in vivo gene therapy may be used to administer TRIP1 or TP2  
CC antisense molecules, or DNA constructs may serve to disrupt or  
CC enhance TRIP1 and/or TP2 expression in cells, and to create  
CC dominant negative inhibitors of TRIP1 or TP2.

XX Sequence 2627 AA;

Query Match 39.1%; Score 43; DB 19; Length 2627;

Best Local Similarity 41.2%; Pred. No. 3.3e+02;

Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Oy 1 VPSYFARAHSEVQDI 17

Db 205 mpsyslsigeeevedl 221

:||||: :|||:

ALT 13

AAG28140

ID AAG28140 standard; Protein: 85 AA.

XX

AC AAG28140;

XX

DT 17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 33246.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-0301439.

XX

PR 25-FEB-1999; 99US-0121825.

PR

PR 05-MAR-1999; 99US-0123180.

PR

PR 09-MAR-1999; 99US-0123548.

PR

PR 23-MAR-1999; 99US-0125788.

PR

PR 25-MAR-1999; 99US-0126264.

PR

PR 29-MAR-1999; 99US-0126785.

PR

PR 01-APR-1999; 99US-0127462.

PR

PR 06-APR-1999; 99US-0128234.

PR

PR 08-APR-1999; 99US-0128714.

PR

PR 16-APR-1999; 99US-0129845.

PR

PR 19-APR-1999; 99US-0130077.

PR

PR 21-APR-1999; 99US-0130449.

PR

PR 23-APR-1999; 99US-0130510.

PR

PR 23-APR-1999; 99US-0130891.

PR

PR 28-APR-1999; 99US-0131449.

PR

PR 30-APR-1999; 99US-0132048.

PR

PR 30-APR-1999; 99US-0132407.

PR

PR 04-MAY-1999; 99US-0132484.

PR

PR 05-MAY-1999; 99US-0132485.

PR

PR 06-MAY-1999; 99US-0132486.

PR

PR 06-MAY-1999; 99US-0132487.

PR

PR 07-MAY-1999; 99US-0132863.

PR

PR 11-MAY-1999; 99US-0134256.

PR

PR 14-MAY-1999; 99US-0134218.

PR

PR 14-MAY-1999; 99US-0134219.

PR

PR 14-MAY-1999; 99US-0134221.

PR

PR 14-MAY-1999; 99US-0134370.

PR

PR 18-MAY-1999; 99US-0134768.

PR

PR 19-MAY-1999; 99US-0134941.

PR

PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144086.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144321.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.

```
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 04-AUG-1999; 99US-0147303.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159284.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.

PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 38.2%; Score 42; DB 21; Length 85;
Best Local Similarity 38.1%; Pred. No. 8.4;
Matches 8; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 1 VPSYSFARAHSEVQDIIRDI 21
Db 22 ipssilarahehdgdeirsv 42
: || |||| : || :

RESULT 14
AAG28139
ID AAG28139 standard; Protein: 91 AA.
XX
AC AAG28139;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 33245.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 200EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
```

PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147302.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match

38.2%;

Score 42; DB 21; Length 91;

Best Local Similarity 38.1%; Pred. No. 9.1;  
Matches 8; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Best Local Similarity 30.0%; Pred. No. 51;  
Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 1 VPSYSFARAHDSVQDIIRD 21  
:|:| |::: :|:|  
Db 28 ipssilarahendgdeefrsv 48

QY 1 VPSYSFARAHDSVQDIIRD 20  
|:| :|:| :|:|  
Db 321 vdytggtgkhdkeilnivke 340

RESULT 15

AAW97743  
ID AAW97743 standard; Protein; 392 AA.

XX AC AAW97743;

XX DT 21-MAY-1999 (first entry)

XX DE Soybean S-adenosylmethionine synthetase.

XX KW S-Adenosylmethionine synthetase; soybean; amino acid; lysine;  
XX KW threonine; methionine; cysteine; isoleucine; transgenic plant;  
XX KW crop improvement; food; feedstuff.

XX OS Glycine max.

XX PN WO9855601-A2.

XX PD 10-DEC-1998.

XX PF 05-JUN-1998; 98WO-US11692.

XX PR 12-JUN-1997; 97US-0049443.

XX PR 06-JUN-1997; 97US-0048771.

XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.

XX PI Abeil LM, Allen SM, Falco SC, Hitz WD, Kinney AJ;

XX PI Rafalski JA, Thorpe CJ;

XX DR WPI: 1999-070263/06.

XX DR N-PSDB; AAX07184.

XX PT New plant amino acid biosynthetic enzymes, DNA and chimeric genes -

XX PT encode: dihydropicolinate reductase; diaminopimelate epimerase;

XX PT threonine synthase; threonine deaminase; S-adenosylmethionine

XX PT synthetase

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

XX PT

Example 7; Page 67-68; 98pp; English.

CC This is the amino acid sequence of a full-length soybean  
CC S-adenosylmethionine synthetase, as deduced from a cDNA clone (see  
CC AAX07184), designated s2.12b06, obtained from a soybean seed  
CC cDNA library. The soybean enzyme shows sequence similarity to the  
CC tomato enzyme. The invention relates to new isolated nucleic acid  
CC fragments (see AAX07168-85) encoding plant enzymes (see AAW97727-44)  
CC that catalyze steps in the biosynthesis of lysine, threonine,  
CC methionine, cysteine and isoleucine from aspartate, the enzyme  
CC being selected from dihydropicolinate reductase,  
CC diaminopimelate epimerase, threonine synthase, threonine deaminase  
CC or S-adenosylmethionine synthetase. The invention also relates to  
CC the construction of a chimeric gene encoding all or a portion of  
CC the biosynthetic pathway enzyme, in sense or antisense orientation,  
CC where expression of the chimeric gene results in production of  
CC altered levels of the enzyme in a transformed host cell.  
CC Overexpression or reduction of expression of genes encoding the  
CC amino acid biosynthetic pathway enzymes in crop plants such as  
CC corn, soybean and wheat can be used to alter levels of the amino  
CC acids in human food and animal feed. Transformed host cells can  
CC also be used to identify compounds that inhibit one of the enzymes.

XX Sequence 392 AA;

A; Molecule type: DNA  
A; Residues: 1-1375 <UED>  
A; Experimental source: GS-5  
R; Shiroza, T.; Ueda, S.; Kuramitsu, H. K.  
J. Bacteriol 169. 4263-4270. 1987

A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.

A:Reference number: A33135; MUID:87308013

A:Accession: C33135

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-349 <SHI>

A:Cross-references: GB:M17361

C:Genetics:

A:Gene: gtfC

C:Function:

A:Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans

C:Superfamily: cpl repeat homology

C:Keywords: duplication; glycosyltransferase; hexosyltransferase

F:1-34/Domain: signal sequence #status predicted <SIG>

F:35-1375/Product: glycosyltransferase #status predicted <MAT>

F:1126-1145/Domain: cpl repeat homology <CP1>

F:1253-1272/Domain: cpl repeat homology <CP2>

F:318-1337/Domain: cpl repeat homology <CP3>

Query Match 90.9%; Score 100; DB 2; Length 1375;

Best Local Similarity 90.9%; Pred. No. 1.8e-07;

Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VPSYSFARHDSVQDLIRDI 22

Db 578 VPSYSFARHDSVQDLIRNII 599

RESULT 3

B33135

gtfB protein precursor - Streptococcus mutans

C:Species: Streptococcus mutans

C:Date: 23-Oct-1990 #sequence\_revision 23-Oct-1990 #text\_change 15-Oct-1999

C:Accession: B33135; A33128

R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

J. Bacteriol. 169, 4263-4270, 1987

A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.

A:Reference number: A33135; MUID:87308013

A:Accession: B33135

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1475 <SHI>

A:Cross-references: GB:M17361; NID:gl53639; PIDN:AAA88588.1; PID:gl53640

R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

submitted to the Protein Sequence Database, September 1990

Reference number: A33128

Accession: A33128

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-171,173-641,'N',643-1475 <SH2>

A:Experimental source: strain GS-5

C:Superfamily: cpl repeat homology

F:1096-1115/Domain: cpl repeat homology <CP1>

F:1224-1243/Domain: cpl repeat homology <CP2>

F:1289-1308/Domain: cpl repeat homology <CP3>

F:1354-1373/Domain: cpl repeat homology <CP4>

F:1419-1438/Domain: cpl repeat homology <CP5>

Query Match

Best Local Similarity 90.0%; Score 99; DB 2; Length 1475;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 VPSYSFARHDSVQDLIRDI 22

Db 552 VPSYSFARHDSVQDLIRNII 573

RESULT 4

JC5473

dextranucrase (EC 2.4.1.5) - Leuconostoc mesenteroides

C:Species: Leuconostoc mesenteroides

C:Date: 07-Jul-1997 #sequence\_revision 29-Aug-1997 #text\_change 29-Aug-1997

C:Accession: JC5473

R:Monchois, V.; Willemot, R.M.; Renaud-Simeon, M.; Croux, C.; Monsan, P.

Gene 182, 23-32, 1996

A:Title: Cloning and sequencing of a gene coding for a novel dextranucrase from Leuc

A:Reference number: JC5473; MUID:97136686

A:Accession: JC5473

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1290 <MON>

A:Cross-references: GB:U38181

C:Comment: This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose

C:Genetics:

A:Gene: dsrA

C:Keywords: glycosyltransferase; hexosyltransferase

F:78-870/Domain: catalytic #status predicted <CAT>

F:922-1290/Domain: glucan-binding #status predicted <GCB>

Query Match 75.5%; Score 83; DB 2; Length 1290;

Best Local Similarity 76.2%; Pred. No. 8.5e-05;

Matches 16; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 PSYSFARHDSVQDLIRDI 22

Db 388 PNYSFIRAHDSVQTIADII 408

RESULT 5

T31098

probable dextranucrase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides

C:Species: Leuconostoc mesenteroides

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000

C:Accession: T31098

R:Monchois, V.; Renaud-Simeon, M.; Monsan, P.; Willemot, R.M.

FEMS Microbiol. Lett. 159, 307-315, 1998

A:Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase

A:Reference number: Z20981; MUID:98164374

A:Accession: T31098

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1508 <MON>

A:Cross-references: EMBL:AF030129; NID:g2766611; PID:g2766612; PIDN:AAB95453.1

A:Experimental source: strain NRRL B-1299

C:Genetics:

A:Gene: dsrB

C:Function:

A:Description: produces dextran composed only of alpha(1-6) glucosidic bonds

C:Keywords: glycosyltransferase; hexosyltransferase

Query Match

Best Local Similarity 72.7%; Score 80; DB 2; Length 1508;

Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 VPSYSFARHDSVQDLIRDI 22

Db 634 IPNYSFVRHDSVQTVIAQII 655

RESULT 6

A41483

glucosyltransferase (EC 2.4.1.1.) gtfS precursor - Streptococcus sobrinus

C:Species: Streptococcus sobrinus

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 15-Oct-1999

C:Accession: A41483

R:Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.

Infect. Immun. 58, 2452-2458, 1990

A:Title: Analysis of the Streptococcus downei gtfS gene, which specifies a glucosyltr

A:Reference number: A41483; MUID:90316665

A:Accession: A41483

A:Molecule type: DNA

A:Residues: 1-1365 <GIL>



```

RESULT 11
A45866
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C:date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text change

```

C;Accession: A45866  
R;Honda, O.; Kato, C.; Kuramitsu, H. K.  
J. Gen. Microbiol. 136, 2099-2105, 1990  
A;Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyl  
A;Reference number: A45866; MUID:91100958  
A;Accession: A45866  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1431 <HON>  
A;Cross-references: GB:M29296  
C;Superfamily: cpl repeat homology  
C;Keywords: glycosyltransferase; hexosyltransferase  
F:181-201/Domain: cpl repeat homology <CP1>  
F:1127-1146/Domain: cpl repeat homology <CP2>  
F:1192-1211/Domain: cpl repeat homology <CP3>  
F:1257-1276/Domain: cpl repeat homology <CP4>  
F:1277-1297/Domain: cpl repeat homology <CP5>  
F:1321-1340/Domain: cpl repeat homology <CP8>  
F:141-1361/Domain: cpl repeat homology <CP6>  
F:1385-1404/Domain: cpl repeat homology <CP7>  
Query Match 56.4%; Score 62; DB 2; Length 1431;  
Best Local Similarity 65.0%; Pred. No. 0.21;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
QY 3 SYSFARHDSVQDLIRDI 22  
Db 576 NIIIRAHDSVQIVIAKII 595  
RESULT 12  
T30858  
glucosyltransferase - Streptococcus salivarius  
C;Species: Streptococcus salivarius  
C;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C;Accession: T30858  
R;Simpson, C.L.; Giffard, P.M.; Jacques, N.A.  
Infect. Immun. 63, 609-621, 1995  
A;Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pri  
A;Reference number: 220909; MUID:951212197  
A;Accession: T30858  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1577 <SIM>  
A;Cross-references: EMBL:L35928; NID:g662380; PID:g662381; PIDN:AACA41413.1  
C;Genetics:  
C;Superfamily: gtfm  
Query Match 56.4%; Score 62; DB 2; Length 1577;  
Best Local Similarity 60.0%; Pred. No. 0.23;  
Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
QY 3 SYSFARHDSVQDLIRDI 22  
Db 661 NYIFVRAHDSVQAVLANII 680  
RESULT 13  
S75540  
hypothetical protein sll1219 - Synecocystis sp. (strain PCC 6803)  
C;Species: Synecocystis sp.  
A;Variety: PCC 6803  
C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
C;Accession: S75540  
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis  
s.  
A;Reference number: S74322; MUID:97061201  
A;Accession: S75540

A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-108 <KAN>  
A;Cross-references: EMBL:D90911; GB:AB001339; NID:gl653083; PIDN:BAA18101.1; PID:gl65  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C;Superfamily: Synecocystis hypothetical protein sll1219  
Query Match 44.5%; Score 49; DB 2; Length 108;  
Best Local Similarity 38.9%; Pred. No. 1.2;  
Matches 7; Conservative 8; Mismatches 3; Indels 0; Gaps 0;  
QY 5 SFARHDSVQDLIRDI 22  
Db 34 NWARHDSQSLDDVVQEIL 51  
RESULT 14  
G82455  
hypothetical protein VCA0471 [Imported] - Vibrio cholerae (strain N16961 serogroup O1  
C;Species: Vibrio cholerae  
C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C;Accession: G82455  
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.  
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers  
L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A;Reference number: A82035; MUID:20406833  
A;Accession: G82455  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-51 <HEI>  
A;Cross-references: GB:AE004379; GB:AE003853; NID:g9657865; PIDN:AAF96375.1; GSPDB:GN  
A;Experimental source: serogroup O1; strain N16961; biotype EI Tor  
C;Genetics:  
A;Gene: VCA0471  
A;Map position: 2  
Query Match 42.7%; Score 47; DB 2; Length 51;  
Best Local Similarity 50.0%; Pred. No. 1;  
Matches 11; Conservative 6; Mismatches 3; Indels 2; Gaps 1;  
QY 1 VPSYSFARHDSVQDLIRDI 22  
Db 15 VPAYSATR--NSEIRDLKEII 34  
RESULT 15  
E83788  
galactose-1-phosphate uridylyltransferase galT [Imported] - Bacillus halodurans (strain  
C;Species: Bacillus halodurans  
C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 08-Dec-2000  
C;Accession: E83788  
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H  
Nucleic Acids Res. 28, 4317-4331, 2000  
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a  
A;Reference number: A83650; MUID:20263314  
A;Accession: E83788  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-508 <STO>  
A;Cross-references: GB:AF001510; GB:BA000004; NID:gl0173440; PIDN:BA04828.1; GSPDB:G  
A;Experimental source: strain C-125  
C;Genetics:  
A;Gene: galT  
C;Superfamily: galactose-1-phosphate uridylyl transferase  
Query Match 42.7%; Score 47; DB 2; Length 508;  
Best Local Similarity 56.2%; Pred. No. 15;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 7 ARAHSEVQDLIRDII 22  
| | | | | | | | | |  
Db 460 AHLHDSNVMDILRDEI 475

Search completed: March 27, 2002, 14:01:24  
Job time: 486 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:27:07 ; Search time 53.4 seconds  
(without alignments)  
15.105 Million cell updates/sec

Title:.. US-09-290-049a-17  
Perfect score: 110  
Sequence: 1 VPSYFARAHDSVQDLIRDI 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues  
Total number of hits satisfying chosen parameters: 100059

Maximum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	110	100.0	1597	1	GTF1_STRDO
2	108	98.2	1592	1	GTF2_STRDO
3	105	95.5	1476	1	GTFB_STRMU
4	100	90.9	1375	1	GTFB_STRMU
5	70	63.6	1365	1	GTFB_STRDO
6	62	56.4	1462	1	GTFD_STRMU
7	47	42.7	508	1	GAL7_BACHD
8	45	40.9	1058	1	POL3_DROME
9	44	40.0	330	1	YH75_ARCFU
10	44	40.0	540	1	ANPC_HUMAN
11	44	40.0	6359	1	BACC_BACLI
12	42.5	38.6	540	1	HXTD_YEAST
13	42	38.2	196	1	MAD2_YEAST
14	42	38.2	280	1	YIBQ_HAEIN
15	42	38.2	429	1	TF3A_YEAST
16	42	38.2	1046	1	RPOC_WEIHE
17	42	38.2	1389	1	PRAX_RAT
18	42	38.2	1835	1	DURL_YEAST
19	41	37.3	295	1	Y326_MYGE
20	41	37.3	306	1	FMRF_LYMTS
21	41	37.3	526	1	MIFC_MOUSE
22	41	37.3	547	1	FUMA_ECOLI
23	41	37.3	933	1	ODOL_ECOLI
24	41	37.3	953	1	COPB_HUMAN
25	41	37.3	953	1	COPB_MOUSE
26	41	37.3	953	1	COPB_RAT
27	40.5	36.8	196	1	YP71_CAEEL
28	40.5	36.8	395	1	VORA_PYRAB
29	40.5	36.8	484	1	SYE_MYCPN
30	40	36.4	124	1	RBS_PSEHY
31	40	36.4	314	1	TOPI_SFVKA
32	40	36.4	344	1	M12D_BACSU
33	40	36.4	387	1	REQN_MOUSE

34	40	36.4	397	1	REQN_RAT	P56163	rattus norv
35	40	36.4	495	1	Y892_MYCTU	Q10532	mycobacteri
36	40	36.4	537	1	ANPC_BOVIN	P10730	bos taurus
37	40	36.4	603	1	PRIM_BACSU	P03096	bacillus su
38	40	36.4	609	1	YLI5_MYCLE	P46509	mycobacteri
39	40	36.4	919	1	Y893_HUMAN	O94967	homo sapien
40	40	36.4	1068	1	P11A_BOVIN	P32871	bos taurus
41	40	36.4	1068	1	P11A_HUMAN	P42336	homo sapien
42	40	36.4	1068	1	P11A_MOUSE	P42337	mus musculus
43	40	36.4	1452	1	VGL2_FIPV	P10033	feline infe
44	39.5	35.9	759	1	SCT1_YEAST	P32784	saccharomyc
45	39	35.5	109	1	CYC6_CYACA	Q9T1W1	cyanidium c

ALIGNMENTS

RESULT 1							
ID	GTF1_STRDO	STANDARD;	PRT;	1597	AA.		
AC	P11001:						
DT	01-JUL-1989	(Rel. 11, Created)					
DT	01-JUL-1989	(Rel. 11, Last sequence update)					
DT	15-DEC-1998	(Rel. 37, Last annotation update)					
DE	GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)						
DE	(SUCROSE 6-GLUCOSYLTRANSFERASE).						
GN	GTFI.						
OS	Streptococcus downei (Streptococcus sobrinus).						
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;						
OC	Streptococcus.						
OX	NCBI_TaxID=1317;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RC	STRAIN=MFE28;						
RA	MEDLINE=87308014; PubMed=3040686;						
RA	Feiretti J.J., Gilpin M.L., Russell R.R.B.;						
RT	"Nucleotide sequence of a glucosyltransferase gene from Streptococcus						
RT	sobrinus Mfe28."						
RL	J. Bacteriol. 169:4271-4278(1987).						
CC	- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT						
CC	TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE						
CC	OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE						
CC	AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.						
CC	- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =						
CC	D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).						
CC	- SUBCELLULAR LOCATION: SECRETED.						
CC	- DISEASE: DENTAL CARIES.						
CC	- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA						
CC	1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES						
CC	WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH						
CC	FORMS OF GLUCANS.						
CC	- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-						
CC	BINDING PROTEIN FROM S. MUTANS.						

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

CC	EMBL; M17391; AAC63063.1; -						
DR	InterPro; IPR002479; CW.binding.						
DR	InterPro; IPR003318; Glyco_hydro_70.						
DR	Pfam; PF01473; CW_binding_1; 19.						
DR	Pfam; PF02324; Glyco_hydro_70; 1.						
KW	Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.						
FT	SIGNAL 1 38						
FT	CHAIN 39 1597						
FT	DOMAIN 39 1050						
FT	DOMAIN 1099 1597						
FT	DOMAIN 1099 1597						
FT	DOMAIN 1.25 A, 2 B, AND 5 AC REPEATS.						





```

RESULT 6
GTFD_STRMU STANDARD; PRT: 1462 AA.
ID GTFD_STRMU STANDARD; PRT: 1462 AA.
AC P49331; O69383; O69386; O69389; O69392; O69398;
DT 01-FEB-1996 (Rel. 33, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTFD.
OS Streptococcus mutans.
OS Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=91100958; PubMed=21486600;
RZ Honda O., Kato C., Kuramitsu H.K.;
RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding
RL the glucosyltransferase-S enzyme.";
RL J. Gen. Microbiol. 136:2099-2105(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MT4239, MT4245, MT4251, MT4467, AND MT8148;
RX MEDLINE=98231643; PubMed=9570124;
RZ Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RL Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
RC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -!- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- DISEASE: DENTAL CARIES.
CC -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S.MUTANS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```



```
CC EMBL; M29296; AAA26895.1; -
DR EMBL; D88653; BAA26103.1; -
DR EMBL; D88656; BAA26107.1; -
DR EMBL; D88659; BAA26111.1; -
DR EMBL; D88662; BAA26115.1; -
DR EMBL; D89979; BAA26121.1; -
DR InterPro; IPR002479; CW_Binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 11.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1 ?
FT CHAIN ?
FT DOMAIN 1232 1423
FT REPEAT 1232 1295
FT REPEAT 1296 1359
FT REPEAT 1360 1423
FT VARIANT 58 58
FT VARIANT 68 68
FT VARIANT 81 81
FT VARIANT 113 113
FT VARIANT 122 122
FT VARIANT 132 132
FT VARIANT 135 135
FT VARIANT 202 202
FT VARIANT 255 255
FT VARIANT 275 275
FT VARIANT 288 288
FT VARIANT 301 301
FT VARIANT 313 313
FT VARIANT 317 317
FT VARIANT 328 328
FT VARIANT 350 350
FT VARIANT 628 633
FT VARIANT 688 688
FT VARIANT 726 732
FT VARIANT 726 730
FT VARIANT 964 964
FT VARIANT 1019 1019
FT VARIANT 1059 1060
FT VARIANT 1060 1060
FT VARIANT 1080 1080
FT VARIANT 1142 1142
FT VARIANT 1198 1198
FT VARIANT 1220 1220
FT VARIANT 1280 1280
FT VARIANT 1282 1282
FT VARIANT 1290 1290
FT VARIANT 1311 1311
FT VARIANT 1403 1403
FT VARIANT 1425 1425
FT VARIANT 1449 1449
FT CONFLICT 1428 1462
FT SEQUENCE 1462 AA; 163512 MW; 5C6541F0DCBDF00 CRC64;
Query Match 56.4%; Score 62; DB 1; Length 1462;
Best Local Similarity 65.0%; Pred. No. 0.13; Indels 5; Gaps 0;
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
```

QY 3 SYSFARHDSVQDLIRDI 22  
: | | | | | | | | | | | |

```
Db 576 NYIFIRAHDSVQTVIAKII 595
RESULT 7
GAL7_BACHD
ID GAL7_BACHD STANDARD; PRT; 508 AA.
AC 09KDV2;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GALACTOSE-1-PHOSPHATE URIDYLTRANSFERASE (EC 2.7.7.10) (GAL-1-P
URIDYLTRANSFERASE).
GN GALT OR BHI109.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=85665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kihara S.,
Horikoshi K.;
RA "Complete genome sequence of the alkaliphilic bacterium Bacillus
halodurans and genomic sequence comparison with Bacillus subtilis.";
RT Nucleic Acids Res. 28:4317-4331(2000).
CC -|- CATALYTIC ACTIVITY: UTP + ALPHA-D-GALACTOSE 1-PHOSPHATE =
UDP-GALACTOSE + PYROPHOSPHATE.
CC -|- PATHWAY: GALACTOSE METABOLISM.
CC -|- SIMILARITY: BELONGS TO THE GALACTOSE-1-PHOSPHATE
URIDYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF001510; BAB04828.1; -
DR InterPro; IPR000880; GalP_UDP_transf.
DR InterPro; IPR000766; GalP_UDP_transf_II.
DR Pfam; PF01087; GalP_UDP_transf. 1.
DR PROSITE; PS01163; GAL_P_UDP_TRANSF_II; 1.
KW Transferase; Nucleotidyltransferase; Galactose metabolism;
KW Complete proteome.
SQ SEQUENCE 508 AA; 57989 MW; 16AF6F607FCEAE2E CRC64;
Query Match 42.7%; Score 47; DB 1; Length 508;
Best Local Similarity 56.2%; Pred. No. 8.6;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
```

QY 7 ARAHDSVQDLIRDI 22  
: | | | | | | | | | | | |

```
Db 460 AHLHDSNVMDILRDEI 475
RESULT 8
POL3_DROME
ID POL3_DROME STANDARD; PRT; 1058 AA.
AC P04323;
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RETROVIRUS-RELATED POL POLYPROTEIN FROM TRANSPOSON 17.6 [CONTAINS:
DE PROTEASE (EC 3.4.23.-); REVERSE TRANSCRIPTASE (EC 2.7.7.49);
DE ENDONUCLEASE].
GN POL.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
```

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85061628; PubMed=6209583;  
 RA Saigo K., Kugimiya W., Matsuo Y., Inouye S., Yoshioka K., Yuki S.;  
 RT "Identification of the coding sequence for a reverse  
 RT transcriptase-like enzyme in a transposable genetic element in  
 RT Drosophila melanogaster.";  
 RL Nature 312:659-661(1984).  
 CC -1- MISCELLANEOUS: THE OPEN READING FRAME IS LOCATED IN A COPIA-LIKE  
 CC TRANSPOSABLE ELEMENT CALLED 17.6.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U22.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X01472; CAA25702.1; -  
 DR PIR; A03971; GNFF17.  
 DR HSP; P04585; LRTI.  
 DR MEROPS; A02.052; -  
 DR FlyBase; FBgn000004; 17.6.  
 DR InterPro; IPR001995; Asp\_prot\_retrov.  
 DR InterPro; IPR001969; Asp\_protease.  
 DR InterPro; IPR000477; RVtse.  
 DR InterPro; IPR001584; Rve.  
 DR Pfam; PF00665; rve; 1.  
 DR Pfam; PF00077; ivp; 1.  
 DR Pfam; PF00078; ivt; 1.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 DR Hydrolase; Aspartyl protease; RNA-directed DNA polymerase;  
 KW Endonuclease; Transferase; Polyprotein; Transposable element.  
 FT ACT\_SITE 30 30 PROTEASE (BY SIMILARITY).  
 SQ SEQUENCE 1058 AA; 122697 MW; C893F5C4A7EIF091 CRC64;

Query Match 40.9%; Score 45; DB 1; Length 1058;  
 Best Local Similarity 36.8%; Pred. No. 39;  
 Matches 7; Conservative 8; Mismatches 4; Indels 0; Gaps 0;  
 [1]  
 4 YSFARAHSEVQDLIRDI 22  
 213 YSPQAYEQVESIQDML 231  
 RESULT 9  
 YH75\_ARCFU  
 ID YH75\_ARCFU STANDARD; PRT; 330 AA.  
 AC O28499;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DE HYPOTHETICAL PROTEIN AFI1775.  
 GN AFI1775.  
 OS Archaeoglobus fulgidus.  
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;  
 OC Archaeoglobus.  
 OX NCBI\_TaxID=2234;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 RX MEDLINE=98049343; PubMed=9389475;  
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
 RA Ketchum K.A., Dodson R.J., Winn M., Hickey E.K., Peterson J.D.,  
 RA Richardson D.L., Kurlavag A.R., Graham D.E., Kyrpides N.C.,  
 RA Fleischmann R.F., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,

RA Peterson S., Reich C.I., McNeill L.K., Badger J.H., Glodek A., Zhou L.,  
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
 RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,  
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing archaeon Archaeoglobus fulgidus.";  
 RL Nature 390:364-370(1997).  
 CC -1- SIMILARITY: BELONGS TO THE ATZ/TRZ FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AE000980; AAB89475.1; -  
 DR TIGR; AFI1775; -  
 DR InterPro; IPR002604; ATZ\_TRZ.  
 DR Pfam; PF01685; ATZ\_TRZ; 1.  
 DR Hypothetical protein; Hydrolase; Complete proteome.  
 KW SEQUENCE 330 AA; 37069 MW; 525AD9F7F35A6FB1 CRC64;  
 SQ SEQUENCE 330 AA; 37069 MW; 525AD9F7F35A6FB1 CRC64;  
 Query Match 40.0%; Score 44; DB 1; Length 330;  
 Best Local Similarity 42.1%; Pred. No. 16;  
 Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;  
 QY 3 SYSFARAHSEVQDLIRDI 21  
 DB 152 AYSSARDHDLKMEVREI 170  
 [1]  
 3 SYSFARAHSEVQDLIRDI 21  
 152 AYSSARDHDLKMEVREI 170  
 RESULT 10  
 ANPC\_HUMAN  
 ID ANPC\_HUMAN STANDARD; PRT; 540 AA.  
 AC P17342;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE ATRIAL NATRIURETIC PEPTIDE CLEARANCE RECEPTOR PRECURSOR (ANP-C)  
 DE (ANPRC) (NPR-C) (ATRIAL NATRIURETIC PEPTIDE C-TYPE RECEPTOR).  
 GN NPR3 OR ANPRC.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE=Kidney;  
 RX MEDLINE=90287735; PubMed=2162522;  
 RA Lowe D.G., Camerato T.R., Goeddel D.V.;  
 RT "cDNA sequence of the human atrial natriuretic peptide clearance  
 RT receptor.";  
 RL Nucleic Acids Res. 18:3412-3412(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90386656; PubMed=2169733;  
 RA Porter J.G., Arfsten A., Fuller F., Miller J.A., Gregory L.C.,  
 RA Lewicki J.A.;  
 RT "Isolation and functional expression of the human atrial natriuretic  
 RT peptide clearance receptor cDNA.";  
 RL Biochem. Biophys. Res. Commun. 171:796-803(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lens epithelium;  
 RA Rae J.L., Shepard A.R.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: RECEPTOR FOR ATRIAL NATRIURETIC PEPTIDE. DOES NOT HAVE  
 CC GUANYLATE CYCLASE ACTIVITY.

CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -1- MISCELLANEOUS: THERE SEEM TO BE AT LEAST THREE ANP RECEPTORS: TWO  
CC WITH GUANYLATE CYCLASE ACTIVITY (ANP-A AND ANP-B) AND ONE (ANP-C)  
CC WHICH IS PROBABLY RESPONSIBLE FOR THE CLEARANCE OF ANP FROM THE  
CC CIRCULATION WITHOUT A ROLE IN SIGNAL TRANSDUCTION.  
CC -1- SIMILARITY: TO ANP-A AND ANP-B RECEPTORS IN THEIR EXTRACELLULAR  
CC AND TRANSMEMBRANE DOMAINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X52282; CAA36523.1; -  
CC EMBL; M59305; AAA51734.1; -  
CC EMBL; AF025998; AAB88801.1; -  
CC PIR; S10150; S10150.  
CC PIR; A35896; A35896.  
CC MIM; 108962.  
CC  
CC InterPro: IPR001170; ANF\_rcptor.  
CC InterPro: IPR001828; ANF\_receptor.  
CC Pfam; PF01094; ANF\_receptor; 1.  
CC PRINTS; PR00255; NATPEPTIDER.  
CC PROSITE; PS00458; ANF\_RECEPTORS; 1.  
CC Receptor; Glycoprotein; transmembrane; Signal.  
CC SIGNAL 1 18  
CC CHAIN 19 540 ATRIAL NATRIURETIC PEPTIDE CLEARANCE  
CC RECEPTOR.  
CC DOMAIN 19 480 EXTRACELLULAR (POTENTIAL).  
CC TRANSMEM 481 503 POTENTIAL.  
CC DOMAIN 504 540 CYTOPLASMIC (POTENTIAL).  
CC DISULFID 108 136 BY SIMILARITY.  
CC DISULFID 213 261 BY SIMILARITY.  
CC DISULFID 473 473 INTERCHAIN (BY SIMILARITY).  
CC CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CONFLICT 476 476 C -> SG (IN REF. 2).  
CC SEQUENCE 540 AA; 59766 MW; 53EE0020A296D6F5 CRC64;  
  
Query Match 40.0%; Score 44; DB 1; Length 540;  
Best Local Similarity 38.9%; Pred. No. 27;  
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;  
  
4 YSFARHDSVQDLIRDI 21  
||||| |:::|::|:  
234 YSFDTKDLELDIVRNI 251  
  
RESULT 11  
ID BACC\_BACLI STANDARD; PRT; 6359 AA.  
AC Q68008;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE BACITRACIN SYNTHETASE 3 (BA3) [INCLUDES: ATP-DEPENDENT ISOLEUCINE  
DE ADENYLASE (ILEA) (ISOLEUCINE ACTIVASE); ATP-DEPENDENT D-PHENYLALANINE  
DE ADENYLASE (D-PHEA) (D-PHENYLALANINE ACTIVASE); ATP-DEPENDENT HISTIDINE  
DE ADENYLASE (HISA) (HISTIDINE ACTIVASE); ATP-DEPENDENT ASPARAGINE  
DE ADENYLASE (D-ASPA) (D-ASPARTATE ACTIVASE); ATP-DEPENDENT ASPARAGINE  
DE ADENYLASE (ASNA) (ASPARAGINE ACTIVASE); ASPARTATE RACEMASE  
DE (EC 5.1.1.13); PHENYLALANINE RACEMASE [ATP HYDROLYSING]  
DE (EC 5.1.1.11)].  
GN BACC.  
OS Bacillus licheniformis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI\_TaxID=1402;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=ATCC 10716;  
RC MEDLINE=98089193; PubMed=9427658;  
RA Konz D., Kleus A., Schoergerdorfer K., Marahiel M.A.;  
RT "The bacitracin biosynthesis operon of Bacillus licheniformis ATCC  
RT 10716: molecular characterization of three multi-modular peptide  
RT synthetases.";  
RL Chem. Biol. 4: 927-937 (1997).  
CC -1- FUNCTION: INDUCES PEPTIDE SYNTHESIS, ACTIVATES AND INCORPORATES  
CC FIVE AMINO ACIDS, FORMS A THIAZOLINE RING BETWEEN THE FIRST TWO  
CC AMINO ACIDS AND INCORPORATES A D-GLUTAMINE IN THE FOURTH POSITION.  
CC -1- CATALYTIC ACTIVITY: L-ASPARTATE = D-ASPARTATE.  
CC -1- CATALYTIC ACTIVITY: ATP + L-PHENYLALANINE = AMP + PYROPHOSPHATE  
CC + D-PHENYLALANINE.  
CC -1- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTETHEINES  
CC (POTENTIAL).  
CC -1- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF THE CYCLIC PEPTIDE  
CC ANTIBIOTIC BACITRACIN.  
CC -1- SUBUNIT: LARGE MULTIENZYME COMPLEX OF BA1, BA2 AND BA3.  
CC -1- DOMAIN: CONSISTS OF FIVE MODULES WITH TWO EPIMERIZATION DOMAINS IN  
CC THE SECOND AND FOURTH MODULES, AND A PUTATIVE C-TERMINAL  
CC THIOESTERASE DOMAIN. EACH MODULE INCORPORATES ONE AMINO ACID INTO  
CC THE PEPTIDE PRODUCT AND CAN BE FURTHER SUBDIVIDED INTO DOMAINS  
CC RESPONSIBLE FOR SUBSTRATE ADENYLATION, THIOATION, CONDENSATION  
CC (NOT FOR THE INITIATION MODULE), AND EPIMERIZATION (OPTIONAL), AND  
CC N METHYLATION (OPTIONAL).  
CC -1- MISCELLANEOUS: BACITRACIN IS A MIXTURE OF AT LEAST TEN CYCLIC  
CC DODECAPEPTIDES, THAT DIFFER BY ONE OR TWO AMINO ACIDS. THE MOST  
CC ABUNDANT IS BACITRACIN A, A BRANCHED CYCLIC DODECAPEPTIDE. IT  
CC CONTAINS AN N-TERMINAL LINEAR PENTAPEPTIDE MOIETY (ILE-CYS-LEU-D-  
CC GLU-ILE) WITH AN ISOLEUCINE-CYSTEINE THIAZOLINE CONDENSATION  
CC PRODUCT AND A C-TERMINAL HEPTAPEPTIDE RING (LYS-D-ORN-ILE-D-PHE-  
CC HIS-D-ASP-ASN), IN WHICH THE FREE ALPHA-CARBOXY GROUP OF THE C-  
CC TERMINAL ASN IS BOUND TO THE EPSILON-AMINO GROUP OF LYSINE. IT  
CC CONTAINS FOUR AMINO ACIDS IN THE D-CONFIGURATION (GLU-4, ORN-7,  
CC PHE-9, AND ASP-11).  
CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME  
CC FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF007865; AAC06348.1; -  
CC InterPro: IPR000873; AMP-bind.  
CC InterPro: IPR001242; DUF4.  
CC InterPro: IPR003880; Phosphopant\_attach.  
CC InterPro: IPR001031; Thioesterase.  
CC Pfam; PF00501; AMP-binding; 5.  
CC Pfam; PF00668; Condensation; 7.  
CC Pfam; PF00550; pp-binding; 5.  
CC Pfam; PF00975; Thioesterase; 1.  
CC PRINTS; PR00154; AMPBINDING.  
CC PROSITE; PS00012; PHOSPHOPANTETHEINE; 4.  
CC PROSITE; PS00455; AMP\_BINDING; 5.  
CC PROSITE; PS00075; ACP\_DOMAIN; 5.  
CC  
CC Ligase; Isomerase; Hydrolase; Antibiotic biosynthesis;  
CC Phosphopantetheine; Multifunctional; Repeat.  
CC REPEAT 461 1034 DOMAIN 1 (ISOLEUCINE-ACTIVATING).  
CC REPEAT 1517 2064 DOMAIN 2 (D-PHENYLALANINE-ACTIVATING).  
CC REPEAT 2999 3570 DOMAIN 3 (HISTIDINE-ACTIVATING).  
CC REPEAT 4047 4612 DOMAIN 4 (D-ASPARTIC ACID-ACTIVATING).  
CC REPEAT 5549 6129 DOMAIN 5 (ASPARAGINE-ACTIVATING).  
CC DOMAIN 966 1034 ACYL CARRIER (ACP) 1.  
CC DOMAIN 1998 2064 ACYL CARRIER (ACP) 2.  
CC DOMAIN 3502 3570 ACYL CARRIER (ACP) 3.  
CC DOMAIN 4544 4612 ACYL CARRIER (ACP) 4.



Query Match 38.2% Score 42; DB 1; Length 196;  
 Best Local Similarity 46.7% Pred. No. 18;  
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Oy 8 RAHDSVODLIRII 22  
 Db 49 KTHDELKDYRKIL 63

## RESULT 14

YIBQ\_HAEIN STANDARD; PRT; 280 AA.  
 AC P44863;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PROTEIN HI0755 PRECURSOR.  
 GN HI0755.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OX Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RD / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus  
 RT influenzae Rd.";  
 RL Science 269:496-512(1995).  
 RN [2]  
 RP IDENTIFICATION BY MASS SPECTROMETRY.  
 RX MEDLINE=20137488; PubMed=10675023;  
 RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,  
 RA Gray C., Fountoulakis M.;  
 RT "Two-dimensional map of the proteome of Haemophilus influenzae.";  
 RL Electrophoresis 21:411-429(2000).  
 CC -1- SIMILARITY: STRONG, TO E.COLI YIBQ.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; U32759; AAC22414.1; ALT\_INIT.  
 CC TIGR; HI0755; -  
 CC SIGNAL; Complete proteome.  
 FT CHAIN 1 26 POTENTIAL.  
 FT CHAIN 27 280 PROTEIN HI0755.  
 SQ SEQUENCE 280 AA; 30740 MW; C739E021613B53B8 CRC64;

Query Match 38.2% Score 42; DB 1; Length 280;  
 Best Local Similarity 36.4% Pred. No. 27;  
 Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Oy 1 VPSYSFAHDSVEQDLIRII 22  
 Db 59 IPAPYARARNQAKSGRDIL 80

## RESULT 15

TF3A\_YEAST STANDARD; PRT; 429 AA.  
 AC P39933;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE TRANSCRIPTION FACTOR IIIA (TFIIIA).  
 GN TFC2 OR P2F1 OR TFIIIA OR YPR186C OR P9677.9.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92147684; PubMed=1737784;  
 RA Archambault J., Milne C.A., Schappert K.T., Baum B., Friesen J.D.,  
 RA Segall J.;  
 RT "The deduced sequence of the transcription factor TFIIIA from  
 RT Saccharomyces cerevisiae reveals extensive divergence from Xenopus  
 RT TFIIIA.";  
 RL J. Biol. Chem. 267:3282-3288(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92237295; PubMed=1570325;  
 RA Woychik N.A., Young R.A.;  
 RT "Genes encoding transcription factor IIIA and the RNA polymerase  
 RT common subunit RPB6 are divergently transcribed in Saccharomycetes  
 RT cerevisiae.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:3999-4003(1992).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
 RA Favell A., Fulton L., Gattung S., Greco T., Kirsten J.,  
 RA Kubaca T., Hallsworth K., Hawkins J., Hillier L., Jier M.,  
 RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,  
 RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,  
 RA Rifken L., Riles L., Taich A., Trevaskis E., Vignati D.,  
 RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterston R.;  
 RL Submitted (APR-1995) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: INTERACTS WITH THE INTERNAL CONTROL REGION (ICR) OF  
 CC APPROXIMATELY 50 BASES WITHIN THE 5S RNA GENES, IS REQUIRED FOR  
 CC CORRECT TRANSCRIPTION OF THESE GENES BY RNA POLYMERASE III. ALSO  
 CC BINDS THE TRANSCRIBED 5S RNA'S.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; M80611; AAB08014.1; -  
 CC EMBL; M90638; -; NOT\_ANNOTATED\_CDS.  
 CC EMBL; U25841; AAB64615.1; -  
 CC PIR; S20050; S20050.  
 CC PIR; A44086; A44086.  
 CC HSP; P08047; I5P2.  
 CC SGD; S0006390; PZF1.  
 CC InterPro; IPR000822; Znf-C2H2.  
 CC Pfam; PF00096; zf-C2H2; 9  
 CC PRINTS; PR00048; ZINC\_FINGER.  
 CC SMART; SM00355; Znf-C2H2; 9.  
 CC PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 8.  
 CC PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 6.  
 CC Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;  
 KW RNA-binding; Repeat; Nuclear protein.  
 FT DOMAIN 24 43 SER-RICH.  
 FT ZN\_FING 49 74 C2H2-TYPE.  
 FT ZN\_FING 80 102 C2H2-TYPE.  
 FT

FT ZN\_FING 108 130 C2H2-TYPE.  
FT ZN\_FING 134 159 C2H2-TYPE.  
FT ZN\_FING 163 186 C2H2-TYPE.  
FT ZN\_FING 194 219 C2H2-TYPE.  
FT ZN\_FING 222 244 C2H2-TYPE.  
FT ZN\_FING 253 277 C2H2-TYPE.  
FT DOMAIN 321 327 ARG/LYS-RICH (BASIC).  
FT ZN\_FING 365 389 C2H2-TYPE.  
SQ SEQUENCE 429 AA; 50027 MW; 209B1E2A20422D9 CRC64;

Query Match 38.2%; Score 42; DB 1; Length 429;  
Best Local Similarity 24.4%; Pred. No. 43;  
Matches 10; Conservative 5; Mismatches 4; Indels 22; Gaps 1;

Oy 2 PSYSFARAH-----SEVQDLIRD 20  
| ||| || |::| |::|  
259 PDMSFSRKHLLTHYGSIHTEEDIPLELKYSIDIQQLVQD 299

Search completed: March 27, 2002, 14:27:08  
Job time: 1650 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:26:09 ; Search time 188.53 Seconds  
(without alignments)  
17.069 Million cell updates/sec

Title: US-09-290-049a-17  
Perfect score: 110  
Sequence: 1 VPSYSFARAHDSVQDLIRDI 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 segs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPREMBL\_17.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	98.2	1590	2	Q55263 streptococc
2	108	98.2	1590	2	Q59983 streptococc
3	105	95.5	1390	2	O69385 streptococc
4	105	95.5	1455	2	O69382 streptococc
5	105	95.5	1455	2	O69391 streptococc
6	105	95.5	1455	2	O69397 streptococc
7	100	90.9	1455	2	O69388 streptococc
8	83	75.5	1290	2	O48756 leuconostoc
9	80	72.7	1477	2	O91466 leuconostoc
10	80	72.7	1508	2	O52224 leuconostoc
11	80	72.7	1508	2	O9EZH5 leuconostoc
12	79	71.8	1016	2	O9LCJ7 leuconostoc
13	79	71.8	1527	2	O9ZAR4 leuconostoc
14	74	67.3	1512	2	O9WJ75 streptococc
15	70	63.6	1338	2	O9WJ74 streptococc
16	70	63.6	1575	2	O9LCH3 streptococc
17	70	63.6	1577	2	O54178 streptococc
18	70	63.6	1599	2	O00599 streptococc
19	68	61.8	1449	2	O68542 streptococc

20	68	61.8	1449	2	Q55264	streptococc
21	68	61.8	2057	2	Q9RE05	leuconostoc
22	65	59.1	1518	2	Q00600	streptococc
23	62	56.4	1577	2	Q55265	streptococc
24	52	47.3	93	2	Q9ZIX9	borrelia bu
25	49	44.5	108	2	P74028	streptococc
26	47	42.7	51	2	Q9RNP7	streptococc
27	47	42.7	361	5	Q9GWP3	vibrio chol
28	47	42.7	597	2	Q9R58	leishmania
29	45	40.9	2627	4	Q99973	ureaplasma
30	44.5	40.5	104	2	Q9X771	homo sapien
31	44	40.0	336	4	Q9Y3S1	listeria mo
32	44	40.0	367	4	Q9H769	homo sapien
33	44	40.0	506	4	Q9NSL5	homo sapien
34	44	40.0	651	5	Q20710	caenorhabdi
35	44	40.0	743	4	Q9NVJ7	homo sapien
36	44	40.0	743	4	Q9NV74	homo sapien
37	44	40.0	743	4	Q9BUN0	homo sapien
38	44	40.0	779	4	Q9H3P4	homo sapien
39	44	40.0	2157	10	Q9AYB5	oryza sativ
40	43.5	39.5	484	2	Q99W75	staphylococ
41	43	39.1	176	2	Q9KD61	staphylococ
42	43	39.1	195	1	Q9YEF4	bacillus ha
43	43	39.1	508	3	Q9P7S7	aeropyrum p
44	43	39.1	570	2	O66690	schizosacch
45	43	39.1	637	1	Q9UZ12	aquifex aeo
						pyrococcus

ALIGNMENTS

RESULT 1  
Q55263 ID Q55263 PRELIMINARY; PRT: 1590 AA.  
AC Q55263;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE GTF-I.  
OS Streptococcus sobrinus,  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1310;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 33478;  
RA Sato S.;  
RL Ann. Kagoshima Univ. Dental School 16:23-29(1996).  
DR EMBL; D63570; BAA09792.1; -  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 15.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1590 AA; 176057 MW; 9DF7A3F2G6E4FD43 CRC64;

Query Match 98.2%; Score 108; DB 2; Length 1590;  
Best Local Similarity 95.5%; Pred. No. 6.2e-08;  
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPSYSFARAHDSVQDLIRDI 22  
Db 548 VPSYSFARAHDSVQDLIRDI 569

RESULT 2  
Q59983 ID Q59983 PRELIMINARY; PRT: 1590 AA.  
AC Q59983;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)





```
Query Match          95.5%; Score 105; DB 2; Length 1455;
Best Local Similarity 95.5%; Pred. No. 1.6e-07;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSFARAHSEVQDLIRDI 22
   ||||| ||||| ||||| ||||| |||||
Db 578 VPSYSFIRAHSEVQDLIRDI 599

RESULT 6
ID O69397 PRELIMINARY; PRT; 1455 AA.
AC O69397;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTFC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT4467;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D89978; BAA26120.1; -.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 9.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR KW Transferase.
SQ SEQUENCE 1455 AA; 162913 MW; A1263427BF24E8E1 CRC64;

Query Match          95.5%; Score 105; DB 2; Length 1455;
Best Local Similarity 95.5%; Pred. No. 1.6e-07;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSFARAHSEVQDLIRDI 22
   ||||| ||||| ||||| ||||| |||||
Db 578 VPSYSFIRAHSEVQDLIRDI 599

RESULT 7
ID O69388 PRELIMINARY; PRT; 1455 AA.
AC O69388;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTFC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT4239;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).

Query Match          95.5%; Score 105; DB 2; Length 1455;
Best Local Similarity 95.5%; Pred. No. 1.6e-07;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSFARAHSEVQDLIRDI 22
   ||||| ||||| ||||| ||||| |||||
Db 578 VPSYSFIRAHSEVQDLIRDI 599

RESULT 8
ID O48756 PRELIMINARY; PRT; 1290 AA.
AC O48756;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DEXTRANSUCRASE.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL B1299;
RX MEDLINE=97136686; PubMed=8982063;
RA Monchois V., Willemot R.M., Remaud-Simeon M., Croux C., Monsan P.;
RA "Cloning and sequencing of a gene coding for a novel dextranucrase
RT from Leuconostoc mesenteroides NRRL B-1299 synthesizing only alpha (1-
RT 6) and alpha (1-3) linkages.";
RL Gene 182:23-32(1996).
DR EMBL; U38181; AAB40875.1; -.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 11.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR KW Transferase.
SQ SEQUENCE 1290 AA; 145590 MW; 3555C2E96B749FAA CRC64;

Query Match          75.5%; Score 83; DB 2; Length 1290;
Best Local Similarity 76.2%; Pred. No. 0.00035;
Matches 16; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 PSYSFARAHSEVQDLIRDI 22
   ||||| ||||| ||||| ||||| |||||
Db 388 PNYSFIRAHSEVQTLIRDI 408

RESULT 9
ID Q9L466 PRELIMINARY; PRT; 1477 AA.
AC Q9L466;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DEXTRANSUCRASE (EC 2.4.1.5).
GN DSRG.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL B-1355;
RA Arguello-Morales M.A., Remaud-Simeon M., Pizzut S., Sarcabal P.,
```

7.

```

RN SEQUENCE FROM N.A.
RP STRAIN=NRRL B-512-F;
RA Bhatnagar R., Singh D.K.S.;
RT "Cloning and Molecular Characterization of Dextranucrase Gene from
RL Leuconostoc mesenteroides NRRL B-512F.";
DR EMBL; U81374; AAD10952.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 16.
DR Pfam; PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1527 AA; 169709 MW; 1DFAFA237C743398 CRC64;

Query Match 71.8%; Score 79; DB 2; Length 1527;
Best Local Similarity 63.6%; Pred. No. 0.0018; Length 1527;
Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHDSVDLIRDI 22
:||||| ||||| ||| :|
652 IPNYSFVRAHDSVQTIVIAQIV 673

RESULT 14
Q9WXJ5 O9WXJ5 PRELIMINARY; PRT; 1512 AA.
AC O9WXJ5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GTF-S.
GN GTF.
OS Streptococcus criceti.
OG Plasmid pAM1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1333;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HS-6;
RA Inoue M., Fukui K., Miyagi A.;
RT "S.cricetus glucosyltransferase(gtfs and gtf) genes.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026123; BAA77237.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 14.
DR Pfam; PF02324; Glyco_hydro_70; 1.
Plasmid.
SEQUENCE 1512 AA; 167145 MW; 4C03D9CBC601FC14 CRC64;

Query Match 67.3%; Score 74; DB 2; Length 1512;
Best Local Similarity 71.4%; Pred. No. 0.01; Length 1512;
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 2 PSYSFARAHDSVDLIRDI 22
||| ||||| ||| :|
560 PSYFVRAHDSVQTIVIAQII 580

RESULT 15
Q9WXJ4 O9WXJ4 PRELIMINARY; PRT; 1338 AA.
AC O9WXJ4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GTF-S.
GN GTF.
OS Streptococcus criceti.
OG Plasmid pAM1.

```

```

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1333;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HS-6;
RA Inoue M., Fukui K., Miyagi A.;
RT "S.cricetus glucosyltransferase(gtfs and gtf) genes.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026123; BAA77236.1; -
DR HSP; P08278; IVJS.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 10.
DR Pfam; PF02324; Glyco_hydro_70; 1.
Plasmid.
KW Plasmid.
SQ SEQUENCE 1338 AA; 148558 MW; 0A90C8E10E15D99B CRC64;

Query Match 63.6%; Score 70; DB 2; Length 1338;
Best Local Similarity 68.2%; Pred. No. 0.037; Length 1338;
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHDSVDLIRDI 22
||| ||||| ||| :|
509 VPNVFIRAHDSVQTRIAKII 530

```

Search completed: March 27, 2002, 14:26:09  
Job time: 1681 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:58:00 ; Search time 198.55 seconds

(without alignments)  
8.208 Million cell updates/sec

Title: US-09-290-049a-18

Perfect score: 110

Sequence: 1 VPVYVFIKSHDSEVQTRIAKII 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A\_Geneseq\_1101.\*  
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	77	70.0	1577	AA91047	Alpha-D-glucosyltr
2	69	62.7	1592	AA91047	Glucosyltransferase
3	63	57.3	2057	AA10667	L. mesenteroides a
4	45	40.9	639	AA117089	Bacillus alkaline
5	44	40.0	583	AA04782	Arabidopsis thalia
6	43	39.1	353	AA37082	Amino acid sequenc
7	42	38.2	3118	AA50362	Human SRCAP. Homo
8	41	37.3	78	AAW21090	Peptide #7524 enco
9	41	37.3	78	AAW37289	Peptide #11326 enc
10	41	37.3	416	AA84536	Amino acid sequenc
11	41	37.3	560	AA842484	Human ORFX ORF2248

12	40	36.4	48	22	AA84625	Amino acid sequenc
13	40	36.4	371	21	AA54076	Enzyme EPS2 involv
14	40	36.4	371	21	AA54076	Amino acid sequenc
15	40	36.4	878	22	AA73265	Yeast trehalose-6-
16	40	36.4	943	21	AA32470	DNA encoding feiln
17	40	36.4	1231	20	AA00219	Enterococcus faeca
18	40	36.4	1265	20	AA00218	Enterococcus faeca
19	40	36.4	2261	22	AA67479	Amino acid sequenc
20	39	35.5	134	21	AA615392	Arabidopsis thalia
21	39	35.5	138	21	AA615391	Arabidopsis thalia
22	39	35.5	193	19	AAW36133	Streptomyces orfl
23	39	35.5	243	20	AA29948	zea mays eif-4E pr
24	39	35.5	245	21	AA54336	Arabidopsis thalia
25	39	35.5	248	21	AA05828	Arabidopsis thalia
26	39	35.5	248	21	AA62363	Arabidopsis thalia
27	39	35.5	248	21	AA34934	Arabidopsis thalia
28	39	35.5	265	21	AA43759	Arabidopsis thalia
29	39	35.5	267	22	AA82207	S. epidermidis ope
30	39	35.5	273	22	AA81536	S. epidermidis ope
31	39	35.5	306	21	AA54335	Arabidopsis thalia
32	39	35.5	309	21	AA05827	Arabidopsis thalia
33	39	35.5	309	21	AA23262	Arabidopsis thalia
34	39	35.5	309	21	AA34933	Arabidopsis thalia
35	39	35.5	326	21	AA05826	Arabidopsis thalia
36	39	35.5	326	21	AA43758	Arabidopsis thalia
37	39	35.5	351	21	AA620875	Arabidopsis thalia
38	39	35.5	351	21	AA53481	Arabidopsis thalia
39	39	35.5	356	21	AA20874	Arabidopsis thalia
40	39	35.5	356	21	AA53480	Arabidopsis thalia
41	39	35.5	375	21	AA54092	Enzyme EPSG involv
42	39	35.5	375	21	AA43794	Amino acid sequenc
43	39	35.5	397	22	AA69077	C glutamicum prote
44	39	35.5	397	22	AA79052	Corynebacterium gl
45	39	35.5	458	21	AA41845	Arabidopsis thalia

#### ALIGNMENTS

RESULT 1

AA91047  
ID AAR91047 standard; Protein; 1577 AA.

XX AAR91047;

XX 22-MAY-1996 (first entry)

XX Alpha-D-glucosyltransferase.

XX Alpha-D-glucosyltransferase; primer-independent; soluble glucan;  
sucrose; transgenic plant; cloning; Escherichia coli;  
phage lambda-cl3; vector; plasmid pSG501; plasmid pSG502;  
gene transfer; crop improvement; storage carbohydrate; pasture;  
feedstuff; senescence; dextran; binder; food; pharmaceutical.

XX Streptococcus salivarius strain ATCC 25975.

XX WO9606173-A1.

XX 29-FEB-1996.

XX 24-AUG-1995; 95WO-AU00527.

XX 24-AUG-1994; 94AU-0007643.

XX (GIFF/) GIFFARD P M.

XX (JACO/) JACQUES N A.

XX (SIMP/) STIMPSON C L.

XX Giffard PM, Jacques NA, Simpson CL;

XX WPI; 1996-151376/15.

DR N-PSDB; AAT13139.

XX Plants contg. new bacterial DNA encoding glucosyl transferase  
 PT activity - retain higher levels of stored carbohydrate(s) in a form  
 PT readily digestible by ruminants  
 XX  
 PS Claim 4; Page 16-20; 31pp; English.  
 XX  
 CC The sequence represents an alpha-D-glucosyltransferase from  
 CC Streptococcus salivarius. The enzyme is primer-independent, and  
 CC produces soluble glucan from sucrose. A gene encoding the enzyme  
 CC may be cloned and expressed in Escherichia coli using a subclone  
 CC of phage lambda-C13, e.g. plasmid pGS501 or plasmid pGS502. The  
 CC DNA may also be expressed in a transgenic plant, to improve the  
 CC level of stored carbohydrate in a pasture plant which normally  
 CC contains low levels, or to prevent degradation of stored carbohydrate  
 CC during plant senescence. Dextran may be isolated from the plant, for  
 CC use as a food binder or pharmaceutical additive. Primer independence  
 CC ensures that the enzyme will be functional in plants. The glucan is  
 CC poorly degraded in plants but easily degraded by bacteria in the rumen  
 CC of grazing livestock.  
 XX  
 SQ Sequence 1577 AA;

Query Match 70.0%; Score 77; DB 17; Length 1577;  
 Best Local Similarity 70.0%; Pred. No. 6.9e-05;  
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 3 NYVFIRAHDSVQTRIAKII 22  
 ||:|||||||:|  
 Db 661 nyifvrahdsqvavlanii 680

RESULT 2  
 AAR32925  
 ID AAR32925 standard; Protein; 1592 AA.

XX AAR32925;  
 AC  
 DT 28-JUN-1993 (first entry)  
 XX  
 DE Glucosyltransferase I.  
 KW GT-1; Streptococcus; dental; caries.  
 XX Streptococcus sobrinus.

JP05023188-A.

XX 02-FEB-1993.

XX 25-JUL-1991; 91JP-0186592.

XX 25-JUL-1991; 91JP-0186592.

XX (FUKU/) FUKUI I.

XX (KATO/) KATO K.

XX WPI; 1993-079449/10.

XX N-PSDB; AAQ37760.

XX DNA sequence glucosyl:transferase-I - comprises Streptococcus

PT sobrinus DNA sequence with at least one nucleotide added or

PT deleted

XX Claim 13; Page 15; 29pp; Japanese.

XX The DNA sequence from Streptococcus sobrinus strain 6715 encodes  
 CC glucosyltransferase-I (and mutants). The DNA was obtd. by treating  
 CC S. sobrinus 6715 with mutanolysin, extracting the chromosomal DNA,  
 CC partially digesting with SauAI and fractionating on agarose gel.  
 CC The 3-5 kbp fragment was ligated into pUC18 and E. coli JM109  
 CC transformed with it. A GT-1 expressing clone was isolated and

CC sequenced. The clone may be used in the development of a drug for  
 CC dental caries.  
 XX  
 SQ Sequence 1592 AA;

Query Match 62.7%; Score 69; DB 14; Length 1592;  
 Best Local Similarity 68.2%; Pred. No. 0.002;  
 Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 VPNVFIRAHDSVQTRIAKII 22  
 ||:|||||||:|  
 Db 548 vpsysfarahdsqvdiirdii 569

RESULT 3

AAB10667  
 ID AAB10667 standard; Protein; 2057 AA.

XX AAB10667;

XX 19-JAN-2001 (first entry)

XX L. mesenteroides alternan sucrose protein.

XX Alternan sucrose; glucosyltransferase; fructose; cosmetic; foodstuff;  
 KW syrup.

XX Leuconostoc mesenteroides.

XX DE19905069-A1.

XX 10-AUG-2000.

XX 08-FEB-1999; 99DE-1005069.

XX 08-FEB-1999; 99DE-1005069.

XX (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.

XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX Kossmann J, Welsh T, Quanz M, Knuth K;

XX WPI; 2000-550294/51.

XX N-PSDB; AAA97904.

XX New nucleic acid encoding recombinant Leuconostoc mesenteroides  
 PT alternan sucrose protein and methods of alternan and fructose  
 PT production

XX Claim 1a; Page 30-36; 64pp; German.

XX This invention describes a novel nucleic acid molecule (I) encoding an  
 CC alternan sucrose (E.C. 2.4.1.140 - an enzyme, that belongs to the  
 CC glucosyltransferase group) The recombinant, purified alternan sucrose  
 CC gene is useful for the fermentative production of alternan (a  
 CC carbohydrate) and/or fructose by secreting the enzyme into a  
 CC saccharose-containing culture medium. Alternatively, the enzyme is  
 CC contacted with a saccharose-containing solution. The alternan and/or  
 CC fructose is then isolated from the medium. Cosmetic products or  
 CC foodstuffs containing alternan can be produced. Recombinant production of  
 CC alternan sucrose is advantageous as it provides a cost effective means of  
 CC producing fructose for high fructose containing syrups, production of  
 CC which previously has been achieved by costly production from maize  
 CC starch. This sequence represents the Leuconostoc mesenteroides alternan  
 CC sucrose protein which is described in the method of the invention.

XX Sequence 2057 AA;

Query Match 57.3%; Score 63; DB 21; Length 2057;

Best Local Similarity 55.0%; Pred. No. 0.033;  
 Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 VPNYVFIRAHDSVQTRIAK 20  
:|||||:|  
Db 757 ipnysfvrahdydaqpik 776

RESULT 4  
AAAY17089  
ID AAY17089 standard; Protein; 639 AA.  
XX  
AC AAY17089;  
XX  
DT 21-JUL-1999 (first entry)  
XX  
DE Bacillus alkaline protease.  
XX  
KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;  
KW washing composition; oxidising agent.  
XX  
OS Bacillus sp.  
XX  
XX WO9918218-A1.  
PD 15-APR-1999.  
XX  
PF 07-OCT-1998; 98WO-JP04528.  
XX  
PR 07-OCT-1997; 97JP-0274570.  
XX  
PA (KAOS )-KAO CORP.  
XX  
XX Hitomi J, Kageyama Y, Kubota H, Nomura M, Okuda M;  
PI Saeki K, Shikata S, Takaiwa M;  
PI WPI; 1999-287736/24.  
DR N-PSDB; AAX37277.  
XX  
XX Alkali protease from Bacillus used in washing powders  
XX  
XX Disclosure; Page 53-58; 71pp; Japanese.  
XX  
CC The invention relates to alkaline proteases produced by strains of  
CC Bacillus. The proteases ability to digest casein is not inhibited by  
CC oleic acid and they have a high stability to oxidising agents. The  
CC alkaline protease of the invention has the following properties: (a) it  
CC is active over the pH range 4-13 and has at least 80% of its optimum  
CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is  
CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)  
CC its ability to digest casein is not inhibited by oleic acid; (e) it has  
CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be  
CC used as enzymes in washing compositions for use in automatic dishwashers  
CC and for washing clothes. The stability to oxidising agents allows the  
CC enzyme to be an effective component of washing compositions including  
CC bleaches. The present sequence represents an alkaline protease.  
XX  
SQ Sequence 639 AA;

Query Match 40.9%; Score 45; DB 20; Length 639;  
Best Local Similarity 28.6%; Pred. No. 15;  
Matches 6; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

QY 1 VPNYVFIRAHDSVQTRIAK 21  
:||:|:|:|:|  
Db 114 ipdyayivegvdqskvrsi 134

RESULT 5  
AAE04782  
ID AAE04782 standard; Protein; 583 AA.  
XX  
AC AAE04782;  
XX

DT 10-SEP-2001 (first entry)  
XX  
DE Arabidopsis thaliana neoxanthin cleavage enzyme, ATNCE1.  
XX  
KW Neoxanthin cleavage enzyme; ATNCE1; abscisic acid; ABA; herbicide;  
KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;  
KW plant growth protectant.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1116794-A2.  
XX  
PD 18-JUL-2001.  
XX  
PF 11-JAN-2001; 2001EP-0300218.  
XX  
PR 13-JAN-2000; 2000JP-0010056.  
PR 11-JAN-2001; 2001JP-0003476.  
XX  
XX (RIKE ) RIKEN KK.  
XX  
XX Iuchi S, Kobayashi M, Shinozaki K;  
PI WPI; 2001-400081/43.  
DR N-PSDB; AAD09394.  
XX  
PT A DNA encoding a protein with a neoxanthin cleavage activity for  
PT producing transgenic plants with improved or decreased stress tolerance  
XX  
PS Claim 3; Fig 10; 101pp; English.  
XX  
CC The invention relates to neoxanthin cleavage enzymes and their  
CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key  
CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.  
CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a  
CC plant when expressed in a plant cell. The invention also relates to  
CC methods for increasing or decreasing stress tolerance in a plant by  
CC introducing the DNA into the plant, and a transgenic plant into which a  
CC neoxanthin cleavage enzyme is introduced. The improvement of stress  
CC tolerance in plants is useful, for example in plant breeding. Neoxanthin  
CC cleavage enzyme genes are useful for producing transgenic plants. An arid  
CC land can be improved by growing transformant weed for several years and  
CC then removing the weed by specifically lowering stress tolerance in the  
CC weed by inducing an inducible promoter. The present sequence is  
CC Arabidopsis thaliana neoxanthin cleavage enzyme, ATNCE1 protein.  
CC The ATNCE1 cDNA is obtained from an Arabidopsis plant-derived cDNA  
CC library using a cDNA of the cPR065 (cowpea Responsive to Dehydration)  
CC gene isolated from cowpea plant as a probe.  
XX  
SQ Sequence 583 AA;

Query Match 40.0%; Score 44; DB 22; Length 583;  
Best Local Similarity 42.1%; Pred. No. 20;  
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 4 YVFIRAHDSVQTRIAKII 22  
:|:|:|:|:|:|  
Db 530 yifchvdeektsetelgii 548

RESULT 6  
AAY37082  
ID AAY37082 standard; Protein; 353 AA.  
XX  
AC AAY37082;  
XX  
DT 07-OCT-1999 (first entry)  
XX  
DE Amino acid sequence of a Chlamydia trachomatis protein.  
XX  
KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;

KW paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;  
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
 KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.  
 XX Chlamydia trachomatis.

OS WO9928475-A2.

XX 10-JUN-1999.

XX 27-NOV-1998; 98WO-IB01939.

XX 04-NOV-1998; 98US-0107077.

XX 28-NOV-1997; 97ER-0015041.

XX 17-DEC-1997; 97ER-0016034.

XX (GEST ) GENSET.

XX Griffals R;

XX WPI; 1999-371125/31.

XX Genome sequence of Chlamydia trachomatis

XX Disclosure; Page 890; 1755pp; English.

XX AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome  
 CC of Chlamydia trachomatis (see AZ01425). The polypeptides can be used as  
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
 CC can also be used to control growth of the microorganism. Chlamydia  
 CC trachomatis is responsible for a large number of diseases, e.g. eye  
 CC diseases such as conventional trachoma, nonendemic trachoma,  
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as  
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,  
 CC perihepatitis, bartholinitis; pneumopathy in breast feeding infants;  
 CC and venereal lymphogranulomatosis. The polypeptides of the invention  
 CC may be of use in treating these diseases.

XX SQ Sequence 353 AA;

Query Match 39.1%; Score 43; DB 20; Length 353;

Best Local Similarity 45.0%; Pred. No. 16;

Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 3 NYVFIHDSVQTRIAKII 22

||||: | | | | | | |

220 nyvlgdrctvrrrrqkll 239

RESULT 7

AAB50362

ID AAB50362 standard; protein; 3118 AA.

XX AAB50362;

XX 12-MAR-2001 (first entry)

XX Human SRCAP.

XX Human; SRCAP; Snf2 related CBP activator protein; antiviral; CREB;

XX CAMP regulatory element; CREB binding protein; CBP; ATPase;

XX transcription activation; DEAD box RNA dependent helicase;

XX adenoviral DBP protein; beta-actin; nuclear receptor; viral infection.

XX Homo sapiens.

XX WO2000073467-A1.

XX 07-DEC-2000.

XX 25-MAY-2000; 2000WO-US14719.

XX

PR 27-MAY-1999; 99US-0136620.

PR 25-MAY-2000; 2000US-0579181.

XX (UYSL-) UNIV SAINT LOUIS.

XX Chrivia J, Yaciuk P;

XX WPI; 2001-061545/07.

XX N-PSDB; AAC89859.

XX Snf2 related CAMP regulatory element (CREB) binding protein (CBP)

XX activator protein, capable of co-activating CREB binding protein,

XX useful for modulating transcription and for affecting viral infection -

XX Claim 6; Page 77-86; 103pp; English.

XX The present sequence is an Snf2 related CREB (CAMP regulatory element)

XX binding protein (CBP) activator protein (SRCAP) polypeptide. It has

XX ATPase activity and is capable of activating transcription. SRCAP

XX polypeptides are useful for activating transcription in a cell, for

XX enhancing CREB (CAMP regulatory element) binding protein (CBP)-mediated

XX activation of transcription in a cell, for treating a patient having a

XX disease involving a function such as insufficient transcription of a

XX gene, e.g. a gene mediated by CBP co-activation, DEAD box RNA dependent

XX helicase, adenoviral DBP protein, beta-actin or a nuclear receptor

XX affected by SRCAP protein. Compounds that modulate SRCAP function, such

XX as antibodies, antisense molecules, polynucleotides or ribozymes, are

XX useful for treating diseases mediated by SRCAP-activated transcription,

XX for example, infection by adenovirus, hepatitis C virus, human

XX immunodeficiency virus type-1, Epstein Barr virus, cytomegalovirus or

XX hepatitis B virus.

XX SQ Sequence 3118 AA;

Query Match 38.2%; Score 42; DB 22; Length 3118;

Best Local Similarity 50.0%; Pred. No. 3.5e+02;

Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 10 HDSEVQTRIAKI 21

||||:|||||

51 heaeletriae 62

RESULT 8

AAM21090

ID AAM21090 standard; Protein; 78 AA.

XX AAM21090;

XX 12-OCT-2001 (first entry)

XX Peptide #7524 encoded by probe for measuring cervical gene expression.

XX Probe; human; microarray; gene expression; cervical epithelial cell;

XX cervical cancer.

XX Homo sapiens.

XX WO200157278-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00670.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX



PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488901/53.  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human cervical epithelial cells -  
XX  
XX  
XX Claim 27; SEQ ID No 25916; 487pp; English.  
XX  
XX The present invention relates to human single exon nucleic acid probes  
CC (SENPs: see AAI10068-AAI28459). The present sequence is a peptide encoded  
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs  
CC can be used to produce a single-exon microarray, which can be used for  
CC measuring human gene expression in a sample derived from human cervical  
CC epithelial cells. By measuring gene expression, the probes are therefore  
CC useful in grading and/or staging of diseases of the cervix, notably  
CC cervical cancer.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
SQ Sequence 78 AA;  
  
Query Match 37.3%; Score 41; DB 22; Length 78;  
Best Local Similarity 38.9%; Pred. No. 6.1;  
Matches 7; Conservative 5; Mismatches 6; Indels 0; Gaps 0;  
  
Qy 4 YVFIRAHDSSEVQTRIARI 21  
:|||||:|||||  
Db 38 fvfikahrfrptgvlk 55  
  
RESULT 9  
AAM37289  
ID AAM37289 standard; Protein: 78 AA.  
XX  
XX AAM37289;  
XX  
XX 17-OCT-2001 (first entry)  
XX  
XX Peptide #11326 encoded by probe for measuring placental gene expression.  
XX Probe: microarray; human; placenta; antenatal diagnosis;  
XX genetic disorder.  
XX  
XX Homo sapiens.  
XX  
XX WO200157272-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US00663.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
XX 03-AUG-2000; 2000US-0632366.  
XX 21-SEP-2000; 2000US-0234687.  
XX 27-SEP-2000; 2000US-0236359.  
XX 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488901/53.  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human placenta -  
XX

PS Claim 27; SEQ ID No 37558; 654pp; English.  
XX  
XX The present invention relates to single exon nucleic acid probes (SENPs:  
CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for producing a microarray for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from human placenta. The probes are useful for antenatal diagnosis of  
CC human genetic disorders.  
XX  
XX Sequence 78 AA;  
  
Query Match 37.3%; Score 41; DB 22; Length 78;  
Best Local Similarity 38.9%; Pred. No. 6.1;  
Matches 7; Conservative 5; Mismatches 6; Indels 0; Gaps 0;  
  
Qy 4 YVFIRAHDSSEVQTRIARI 21  
:|||||:|||||  
Db 38 fvfikahrfrptgvlk 55  
  
RESULT 10  
AAB84536  
ID AAB84536 standard; Protein: 416 AA.  
XX  
XX AAB84536;  
XX  
XX 05-SEP-2001 (first entry)  
XX  
XX Amino acid sequence of clostridial lysine 2,3-aminomutase.  
XX Lysine 2,3-aminomutase; L-lysine; L-beta-lysine; antibiotic.  
XX Clostridium subterminale.  
XX  
XX US6248874-A1.  
XX  
XX 19-JUN-2001.  
XX  
XX 11-JUN-1999; 99US-0330611.  
XX  
XX 24-NOV-1998; 98US-0198942.  
XX  
XX (WISC ) WISCONSIN ALUMNI RES FOUND.  
XX  
XX Frey PA, Ruzicka FJ;  
XX  
XX WPI; 2001-424364/45.  
XX N-PSDB; AAH29114.  
XX  
XX New bacteria lysine 2,3-aminomutase protein and nucleic acid encoding  
PT the protein, useful for preparing L-beta-lysine which is an important  
PT constituent of antibiotics -  
XX  
XX Claim 1; Column 5-6; 44pp; English.  
XX  
XX The present sequence represents a lysine 2,3-aminomutase. The clostridial  
CC enzyme catalyses the reversible isomerisation of L-lysine into  
CC L-beta-lysine. The clostridial enzyme is a hexameric protein of  
CC apparently identical subunits, and contains iron-sulphur clusters,  
CC cobalt and zinc, and pyridoxal 5'-phosphate. The clostridial enzyme  
CC is activated by S-adenosylmethionine. The clostridial enzyme is useful  
CC for preparing L-beta-lysine, which is an important constituent of  
CC antibiotics.  
XX  
XX Sequence 416 AA;  
  
Query Match 37.3%; Score 41; DB 22; Length 416;  
Best Local Similarity 31.8%; Pred. No. 46;  
Matches 7; Conservative 5; Mismatches 10; Indels 0; Gaps 0;  
  
Qy 1 VPNVFIRAHDSSEVQTRIARI 22

```

Db      341 mpnyvisqshdkvllrnfevgvi 362
      :| | | | :| | | : :|

```

RESULT 11

AAB42484  
ID AAB42484 standard; Protein; 560 AA.

XX AAB42484:

XX  
DT 08-FEB-2001 (first entry)

DE Human ORFX 'ORF2248 polypeptide sequence SEQ ID NO:4496. XX

Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; vulnary; antipapillary; antiparkinsonian; nootropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antithyroid; antianaemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive.	KW
--	----

OS Homo sapiens.

AA  
PN  
WO200058473-A2.

05-OCT-2000.  
PD  
XX

31-MAR-2000; 2000WO-US08621.

31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127638.  
PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540/63.  
XXPA (CURA-) CURAGEN CORP.  
XX

Shimkets RA, Leach M  
 PI  
 YY

WPI; 2000-602362/57.  
N-DCDB: 28076603

**XX**

PT useful for treatf

XX  
PS Claim 11; Page 3681-3683; 5507pp; English.

CC AAC74446 to AAC77606 encode the proteins g

which represent the human open reading frames 1 to 3101. The ORF sequences have activities such as: cytostatic; hepatotropic; vulnerable; antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic; anticoagulant; antiarthritic; immunosuppressive; immunostimulant; cardiant; thrombolytic; coagulant; vasotropic; antiadrenergic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORF-associated disorder. The nucleic acids can be used to express ORF proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma,

RESOLUT 13  
AAB73265

Search completed: March 27, 2002, 13:58:01  
Job time: 529 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:59:33 ; Search time 87.3 seconds  
(without alignments)  
5,671 Million cell updates/sec

Title: US-09-290-049a-18

Perfect score: 110

Sequence: 1 VFNVYFIRAHSEVQTRIAKII 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/iaa/backfilesi.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	90	81.8	1430	3	US-09-008-172-2
2	90	81.8	1430	4	US-09-210-361-6
3	80	72.7	1475	3	US-09-007-999-2
4	80	72.7	1475	4	US-09-210-361-2
5	77	70.0	1577	2	US-08-793-824-2
6	76	69.1	1375	4	US-09-210-361-4
7	41	37.3	416	4	US-09-330-611-2
8	40	36.4	48	4	US-09-330-611-21
9	40	36.4	943	3	US-08-911-321-4
10	40	36.4	1041	1	US-08-220-151-4
11	40	36.4	1041	1	US-08-413-118-4
12	40	36.4	1041	3	US-08-473-446-4
13	39	35.5	287	2	US-08-901-306-2
14	39	35.5	287	4	US-09-180-271-2
15	38	34.5	368	4	US-08-709-838-2
16	38	34.5	368	4	US-08-829-839-2
17	38	34.5	615	2	US-08-663-566A-17
18	38	34.5	615	2	US-08-023-610-17
19	38	34.5	615	2	US-08-288-065A-17
20	38	34.5	615	2	US-08-362-240A-17
21	38	34.5	615	5	PCT-US95-10245-17
22	38	34.5	868	2	US-08-907-166-2
23	38	34.5	1162	2	US-08-663-566A-15
24	38	34.5	1162	2	US-08-023-610-15
25	38	34.5	1162	2	US-08-288-065A-15
26	38	34.5	1162	2	US-08-362-240A-15
27	38	34.5	1162	5	PCT-US95-10245-15

28	38	34.5	1464	4	US-08-891-640-2	Sequence 2, Appli
29	37	33.6	553	1	US-08-475-894-2	Sequence 2, Appli
30	37	33.6	553	1	US-08-484-710-2	Sequence 2, Appli
31	37	33.6	553	2	US-08-484-709-2	Sequence 2, Appli
32	37	33.6	553	4	US-08-474-697-2	Sequence 2, Appli
33	37	33.6	936	5	PCT-US93-05944-2	Sequence 3, Appli
34	37	33.6	1118	4	US-09-379-523-3	Sequence 3, Appli
35	36.5	33.2	188	1	US-08-442-063A-39	Sequence 39, Appl
36	36.5	33.2	236	1	US-08-442-063A-42	Sequence 42, Appl
37	36.5	33.2	282	1	US-08-442-063A-45	Sequence 45, Appl
38	36.5	33.2	307	1	US-08-442-063A-48	Sequence 48, Appl
39	36.5	33.2	333	1	US-08-442-063A-27	Sequence 27, Appl
40	36.5	33.2	342	1	US-08-272-919-2	Sequence 2, Appli
41	36.5	33.2	342	1	US-08-619-916-2	Sequence 2, Appli
42	36.5	33.2	342	5	PCT-US95-08542-2	Sequence 2, Appli
43	36.5	33.2	353	6	5340934-4	Patent No. 5340934
44	36.5	33.2	359	1	US-08-303-238-4	Sequence 4, Appli
45	36.5	33.2	359	4	US-08-458-834-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1  
US-09-008-172-2  
; Sequence 2, Application US/09008172  
; Patent No. 6127602  
; GENERAL INFORMATION:  
; APPLICANT: Nicholas, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; FILE REFERENCE: 0358D  
; CURRENT APPLICATION NUMBER: US/09/008,172  
; CURRENT FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-008-172-2

Query Match 81.8%; Score 90; DB 3; Length 1430;  
Best Local Similarity 90.0%; Pred. No. 1.8e-07;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 3 NVVFIHDSSEVQTRIAKII 22  
Db 576 NVVFIHDSSEVQTRIAKII 595  
  
RESULT 2  
US-09-210-361-6  
; Sequence 6, Application US/09210361  
; Patent No. 6284479  
; GENERAL INFORMATION:  
; APPLICANT: Nicholas, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starches and  
; FILE REFERENCE: 0357CR  
; CURRENT APPLICATION NUMBER: US/09/210,361  
; CURRENT FILING DATE: 1998-12-11  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/009,620  
; EARLIER FILING DATE: 1998-01-20  
; EARLIER APPLICATION NUMBER: 08/485,243  
; EARLIER FILING DATE: 1995-06-07

; EARLIER APPLICATION NUMBER: 09/008,172  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 6  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-210-361-6

Query Match 81.8%; Score 90; DB 4; Length 1430;  
Best Local Similarity 90.0%; Pred. No. 1.1e-07;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NVYFIRAHDSVQTRIADII 22  
||:||||||| |||||  
DB 576 NVYFIRAHDSVQTRIADII 595

RESULT 3  
US-09-007-999-2  
; Sequence 2, Application US/09007999  
; Patent No. 6087559  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; FILE REFERENCE: 0356D  
; CURRENT APPLICATION NUMBER: US/09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-007-999-2

Query Match 72.7%; Score 80; DB 3; Length 1475;  
Best Local Similarity 77.3%; Pred. No. 1.1e-05;  
Matches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPNYVFIHDSVQTRIADII 22  
||:||||||| |||||  
DB 552 VPSYFIRAHDSVQDLIADII 573

RESULT 4  
US-09-210-361-2  
; Sequence 2, Application US/09210361  
; Patent No. 6284479  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starches and  
; FILE REFERENCE: 0357CR  
; CURRENT APPLICATION NUMBER: US/09/210,361  
; EARLIER FILING DATE: 1998-12-11  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/009,620  
; EARLIER FILING DATE: 1998-01-20  
; EARLIER APPLICATION NUMBER: 08/485,243  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/008,172

; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-210-361-2

Query Match 72.7%; Score 80; DB 4; Length 1475;  
Best Local Similarity 77.3%; Pred. No. 1.1e-05;  
Matches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPNYVFIHDSVQTRIADII 22  
||:||||||| |||||  
DB 552 VPSYFIRAHDSVQDLIADII 573

RESULT 5  
US-08-793-824-2  
; Sequence 2, Application US/08793824  
; Patent No. 5981838  
; GENERAL INFORMATION:  
; APPLICANT: Simpson, Christine Lynn  
; APPLICANT: Giffard, Phillip Morris  
; APPLICANT: Jacques, Nicholas Anthony  
; TITLE OF INVENTION: Genetic Manipulation of plants to  
; TITLE OF INVENTION: Increase Stored Carbohydrates  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Griffith Hack & Co  
; STREET: Level 8, 168 Walker Street  
; CITY: No. 5981838th Sydney  
; STATE: New South Wales  
; COUNTRY: Australia  
; ZIP: 2060  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/793,824  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: AU PM7643  
; FILING DATE: 24-AUG-1994  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 61 2 9957 5944  
; TELEFAX: 61 2 957 6288  
; TELEX: 26547  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1577 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Streptococcus salivarius  
US-08-793-824-2

Query Match 70.0%; Score 77; DB 2; Length 1577;  
Best Local Similarity 70.0%; Pred. No. 4.1e-05;  
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 NVYFIRAHDSVQTRIADII 22  
||:||||||| |||||

Db 661 NYFVRAHDSVQAVLANII 680

RESULT 6  
US-09-210-361-4  
; Sequence 4, Application US/09210361  
; Patent No. 6284479  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starches and  
; FILE OF INVENTION: Latexes in Paper Manufacture  
; FILE REFERENCE: 0357CR  
; CURRENT APPLICATION NUMBER: US/09/210,361  
; CURRENT FILING DATE: 1998-12-11  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/009,620  
; EARLIER FILING DATE: 1998-01-20  
; EARLIER APPLICATION NUMBER: 08/485,243  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/008,172  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 1375  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-210-361-4

Query Match 69.1%; Score 76; DB 4; Length 1375;  
Best Local Similarity 72.7%; Pred. No. 5.2e-05;  
Matches 16; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 VPNVFIRAHDSVQVTRIAKII 22  
|||||:|||||:|

Db 578 VPSYFIRAHDSVQDLIRNII 599

RESULT 7  
US-09-330-611-2  
; Sequence 2, Application US/09330611  
; Patent No. 6248874  
; GENERAL INFORMATION:  
; APPLICANT: FREY, PERRY A.  
; TITLE OF INVENTION: DNA MOLECULES ENCODING BACTERIAL LYSINE 2,3-AMINOMUTASE  
; FILE REFERENCE: 032026/0476  
; CURRENT APPLICATION NUMBER: US/09/330,611  
; CURRENT FILING DATE: 1999-06-11  
; EARLIER APPLICATION NUMBER: US/09/198,942  
; EARLIER FILING DATE: 1998-11-24  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 416  
; TYPE: PRT  
; ORGANISM: Clostridium subterminale  
US-09-330-611-2

Query Match 37.3%; Score 41; DB 4; Length 416;  
Best Local Similarity 31.8%; Pred. No. 21;  
Matches 7; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

Qy 1 VPNVFIRAHDSVQVTRIAKII 22  
|||||:|||||:|

Db 341 MPNVVISQSHDKVILRNFEVGI 362

RESULT 8  
US-09-330-611-21  
; Sequence 21, Application US/09330611  
; Patent No. 6248874  
; GENERAL INFORMATION:  
; APPLICANT: FREY, PERRY A.  
; TITLE OF INVENTION: DNA MOLECULES ENCODING BACTERIAL LYSINE 2,3-AMINOMUTASE  
; FILE REFERENCE: 032026/0476  
; CURRENT APPLICATION NUMBER: US/09/330,611  
; CURRENT FILING DATE: 1999-06-11  
; EARLIER APPLICATION NUMBER: US/09/198,942  
; EARLIER FILING DATE: 1998-11-24  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 21  
; LENGTH: 48  
; TYPE: PRT  
; ORGANISM: Clostridium subterminale  
US-09-330-611-21

Query Match 36.4%; Score 40; DB 4; Length 48;  
Best Local Similarity 33.3%; Pred. No. 2.5;  
Matches 7; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

Qy 2 PNVVFIRAHDSVQVTRIAKII 22  
|||||:|||||:|

Db 1 PNVVISQSHDKVILRNFEVGI 21

RESULT 9  
US-08-911-321-4  
; Sequence 4, Application US/08911321  
; Patent No. 6010703  
; GENERAL INFORMATION:  
; APPLICANT: Roger K. Maes and Stephen J. Spatz  
; TITLE OF INVENTION: Recombinant Poxvirus  
; TITLE OF INVENTION: Vaccine Against  
; TITLE OF INVENTION: Feline Rhinotracheitis  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ian C. McLeod  
; STREET: 2190 Commons Parkway  
; CITY: Okemos  
; STATE: Michigan  
; COUNTRY: USA  
; ZIP: 48864  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/911,321  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/096,183  
; FILING DATE: July 26, 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ian C. McLeod  
; REGISTRATION NUMBER: 20,931  
; REFERENCE/DOCKET NUMBER: MSU 4.1-166  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (517) 347-4100  
; TELEFAX: (517) 347-4103  
; TELEX: No. 6010703e  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:

LENGTH: 943  
TYPE: Amino Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE: Polypeptide  
DESCRIPTION: Polypeptide  
HYPOTHETICAL: No  
ANTI-SENSE: No  
ORIGINAL SOURCE:  
ORGANISM: Feline herpesvirus-1  
STRAIN: 1  
INDIVIDUAL ISOLATE: C-27  
CELL TYPE: N/A  
FEATURE:  
NAME/KEY:  
LOCATION:  
IDENTIFICATION METHOD: Deduced sequence  
OTHER INFORMATION: gb  
08-911-321-4

Query Match 36.4%; Score 40; DB 3; Length 943;  
Best Local Similarity 50.0%; Pred. No. 83;  
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 4 YVFIRAHDSVQTRIA 19  
| :|:| :|:| :|:|  
Db 545 YDIQAHVNEMLSRIA 560

RESULT 10  
US-08-220-151-4  
; Sequence 4, Application US/08220151  
; Patent No. 5529780  
; GENERAL INFORMATION:  
; APPLICANT: Paolletti, Enzo  
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF  
; CANINE HERPESVIRUS gb, gc AND gd AND USES THEREFOR  
; NUMBER OF SEQUENCES: 91  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Curtis, Morris & Safford  
; STREET: 530 Fifth Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/220,151  
; FILING DATE: 30-MAR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Frommer, William S.  
; REGISTRATION NUMBER: 25,506  
; REFERENCE/DOCKET NUMBER: 454310-2540  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 840-3333  
; TELEFAX: (212) 840-0712  
; TELEX: 425066 CURTMS  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1041 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: N-terminal  
US-08-220-151-4

Query Match 36.4%; Score 40; DB 1; Length 1041;  
Best Local Similarity 50.0%; Pred. No. 93;  
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 4 YVFIRAHDSVQTRIA 19  
| :|:| :|:| :|:|  
Db 641 YDIQAHVNEMLSRIA 656

RESULT 11  
US-08-413-118-4  
; Sequence 4, Application US/08413118  
; Patent No. 5688920  
; GENERAL INFORMATION:  
; APPLICANT: PAOLETTI, ENZO  
; APPLICANT: LIMBACH, KEITH J.  
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF  
; CANINE HERPESVIRUS gb, gc, AND gd AND USES THEREFOR  
; NUMBER OF SEQUENCES: 128  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.  
; STREET: 530 FIFTH AVENUE, 25TH FLOOR  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/413,118  
; FILING DATE: 29-MAR-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/220,151  
; FILING DATE: 30-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FROMMER, WILLIAM S.  
; REGISTRATION NUMBER: 25,506  
; REFERENCE/DOCKET NUMBER: 454310-2670  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 840-3333  
; TELEFAX: (212) 840-0712  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1041 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: N-terminal  
US-08-413-118-4

Query Match 36.4%; Score 40; DB 1; Length 1041;  
Best Local Similarity 50.0%; Pred. No. 93;  
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 4 YVFIRAHDSVQTRIA 19  
| :|:| :|:| :|:|  
Db 641 YDIQAHVNEMLSRIA 656

RESULT 12  
US-08-473-446-4  
; Sequence 4, Application US/08473446  
; Patent No. 6017542  
; GENERAL INFORMATION:  
; APPLICANT: PAOLETTI, ENZO



APPLICANT: LIMBACH, KEITH J.  
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF  
CANINE HERPESVIRUS gB, gC, AND gD AND USES THEREFOR  
NUMBER OF SEQUENCES: 128  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.  
STREET: 530 FIFTH AVENUE, 25TH FLOOR  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/08/473,446  
FILING DATE:  
CLASSIFICATION:  
PRIORITY APPLICATION DATA: 08/413,118  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: FROMMER, WILLIAM S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2670  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1041 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-473-446-4

Query Match 36.4%; Score 40; DB 3; Length 1041;  
Best Local Similarity 50.0%; Pred. No. 93;  
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 4 YVFIKRDSEVQTRIA 19  
I : : : : :  
Db 641 YDVIQAHVNEMLSR 656

RESULT 13  
US-08-901-306-2  
Sequence 2, Application US/08901306  
Patent No. 5955319  
GENERAL INFORMATION:  
APPLICANT: FILIPPINI, Silvia  
APPLICANT: BREME, Umberto  
APPLICANT: COLOMBO, Anna L.  
APPLICANT: LOMOVSKAYA, Natasha  
APPLICANT: FONSTEIN, Leonid  
APPLICANT: OTTEN, Sharee  
TITLE OF INVENTION: PROCESS FOR PREPARING DOXORUBICIN  
FILE REFERENCE: 1615-7003  
CURRENT APPLICATION NUMBER: US/08/901,306  
CURRENT FILING DATE: 1997-07-28  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 287  
TYPE: PRT  
ORGANISM: Streptomyces peucetius  
US-08-901-306-2

Query Match 35.5%; Score 39; DB 2; Length 287;  
Best Local Similarity 36.8%; Pred. No. 31;  
Matches 7; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 2 PNVFIRAHDSVQTRIAK 20  
I : : : : :  
Db 194 PDVLAVSYHFGVTRIGR 212

RESULT 14  
US-09-180-271-2  
Sequence 2, Application US/09180271  
Patent No. 6210930  
GENERAL INFORMATION:  
APPLICANT: FILIPPINI, Silvia  
APPLICANT: LOMOVSKAYA, Natalia  
APPLICANT: FONSTEIN, Leonid  
APPLICANT: COLOMBO, Anna L.  
APPLICANT: HUTCHINSON, C. Richard  
APPLICANT: OTTEN, Sharee L.  
APPLICANT: BREME, Umberto  
TITLE OF INVENTION: PROCESS FOR PREPARING DOXORUBICIN  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nikaido, Marmelstein, Murray & Oram  
STREET: 655 Fifteenth Street N.W. Suite 330  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-5701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/180,271  
FILING DATE: 14-JAN-1999  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: PCT/US98/03938  
FILING DATE: 05-MAR-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/901,306  
FILING DATE: 28-JUL-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/812,412  
FILING DATE: 06-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Kitts, Monica C.  
REGISTRATION NUMBER: 36,105  
REFERENCE/DOCKET NUMBER: P1615-8010  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)638-5000  
TELEFAX: (202)638-4810  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 287 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-180-271-2

Query Match 35.5%; Score 39; DB 4; Length 287;  
Best Local Similarity 36.8%; Pred. No. 31;  
Matches 7; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 2 PNVFIRAHDSVQTRIAK 20  
I : : : : :  
Db 194 PDVLAVSYHFGVTRIGR 212

RESULT 15  
US-08-709-838-2  
; Sequence 2, Application US/08709838  
; Patent No. 6140064  
; GENERAL INFORMATION:  
; APPLICANT: Loetscher, Marcel  
; APPLICANT: Moser, Bernhard  
; TITLE OF INVENTION: IP-10/MIG RECEPTOR DESIGNATED CXCR3  
; TITLE OF INVENTION: NUCLEIC ACIDS, AND METHODS OF USES THEREFOR  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/709,838  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brook Esq., David E.  
; REGISTRATION NUMBER: 22,592  
; REFERENCE/DOCKET NUMBER: TK196-01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 861-6240  
; TELEFAX: (617) 861-9540  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 368 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-709-838-2

Query Match 34.5%; Score 38; DB 4; Length 368;  
Best Local Similarity 38.5%; Pred. No. 62;  
Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;  
QY 1 VPNTVFIRADSE 13  
:|::|: || |  
Db 184 LPDFILSAHDE 196

Search completed: March 27, 2002, 13:59:34  
Job time: 587 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:20:30 ; Search time 1139.61 seconds  
(without alignments)  
5.360 Million cell updates/sec

Title: US-09-290-049a-18  
Perfect score: 110  
Sequence: 1 VPNTVFIRAHDSVQTRIAKII 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 3148936 seqs, 277657034 residues  
Total number of hits satisfying chosen parameters: 3148936

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/2/paa/PCTUS\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/paa/US080\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pep.\*  
7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pep.\*  
8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep.\*  
9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pep.\*  
10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep.\*  
11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pep.\*  
12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pep.\*  
13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep.\*  
14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep.\*  
15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pep.\*  
16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pep.\*  
17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pep.\*  
18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pep.\*  
19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pep.\*  
20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pep.\*  
21: /cgn2\_6/ptodata/2/paa/US097\_COMB.pep.\*  
22: /cgn2\_6/ptodata/2/paa/US098\_COMB.pep.\*  
23: /cgn2\_6/ptodata/2/paa/US099\_COMB.pep.\*  
24: /cgn2\_6/ptodata/2/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	100.0	22	16	US-09-290-049-18
2	90	81.8	22	16	US-09-290-049-16
3	90	81.8	1430	20	US-09-649-885-2
4	90	81.8	1430	21	US-09-740-274-6
5	80	72.7	1475	19	US-09-557-848-2
6	80	72.7	1475	21	US-09-740-274-2
7	76	69.1	22	16	US-09-290-049-15
8	76	69.1	1375	21	US-09-740-274-4
9	70	63.6	22	16	US-09-290-049-17
					Sequence 18, Appl
					Sequence 16, Appl
					Sequence 2, Appli
					Sequence 6, Appli
					Sequence 2, Appli
					Sequence 2, Appli
					Sequence 15, Appl
					Sequence 4, Appli
					Sequence 17, Appl

69	62.7	22	16	US-09-290-049-19	Sequence 19, Appl
63	57.3	2057	18	US-09-499-203-2	Sequence 2, Appli
62.5	56.8	20	16	US-09-290-049-2	Sequence 2, Appli
47	42.7	336	24	US-60-161-932-1239	Sequence 1239, Ap
47	42.7	336	24	US-60-167-217-3108	Sequence 3108, Ap
47	42.7	336	24	US-60-173-464-2558	Sequence 2558, Ap
47	42.7	336	24	US-60-191-637-3060	Sequence 3060, Ap
47	42.7	336	24	US-60-191-681-2460	Sequence 2460, Ap
45.5	41.4	133	16	US-09-270-767-32582	Sequence 32582, A
45.5	41.4	133	16	US-09-270-767-47799	Sequence 47799, A
45.5	41.4	133	16	US-09-270-849B-185935	Sequence 185935, A
45.5	41.4	866	24	US-60-167-217-21484	Sequence 21484, A
45.5	41.4	866	24	US-60-173-464-17477	Sequence 17477, A
45.5	41.4	866	24	US-60-191-637-21353	Sequence 21353, A
45.5	41.4	866	24	US-60-191-681-16832	Sequence 16832, A
45	40.9	639	19	US-09-509-814A-4	Sequence 4, Appli
44	40.0	583	21	US-09-758-269-2	Sequence 2, Appli
44	40.0	737	24	US-60-150-584-711	Sequence 711, App
43	39.1	93	18	US-09-417-507-36660	Sequence 36660, A
43	39.1	355	16	US-09-201-228A-330	Sequence 330, App
43	39.1	424	16	US-09-270-767-43585	Sequence 43585, A
42	38.2	76	21	US-09-758-472-5026	Sequence 5026, Ap
42	38.2	183	15	US-09-107-532-5511	Sequence 5511, Ap
42	38.2	183	15	US-09-107-532A-5511	Sequence 11822, A
42	38.2	472	24	US-60-167-217-11822	Sequence 9579, Ap
42	38.2	472	24	US-60-173-464-9579	Sequence 11840, A
42	38.2	472	24	US-60-191-637-11840	Sequence 9284, Ap
42	38.2	652	17	US-09-328-352-5587	Sequence 5587, Ap
42	38.2	3118	19	US-09-579-181-1	Sequence 1, Appli
41.5	37.7	263	21	US-09-733-089-19824	Sequence 19824, A
41.5	37.7	263	22	US-09-816-660-19824	Sequence 19824, A
41	37.3	78	1	PCT-US01-00663-37558	Sequence 37558, A
41	37.3	78	22	US-09-864-761-41675	Sequence 41675, A
41	37.3	90	16	US-09-248-796-17154	Sequence 17154, A
41	37.3	191	18	US-09-417-507-26018	Sequence 26018, A

ALIGNMENTS

RESULT 1  
US-09-290-049-18  
; Sequence 18, Application US/09290049  
; GENERAL INFORMATION:  
; APPLICANT: Taubman, Daniel J.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: FDC98-01p2A  
; CURRENT APPLICATION NUMBER: US/09/290,049  
; CURRENT FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081,550  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/115,142  
; EARLIER FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 18  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. downei  
US-09-290-049-18

Query Match 100.0%; Score 110; DB 16; Length 22;  
Best Local Similarity 100.0%; Pred. No. 9.1e-12;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPNTVFIRAHDSVQTRIAKII 22  
Db 1 VPNTVFIRAHDSVQTRIAKII 22

RESULT 2  
US-09-290-049-16  
; CURRENT APPLICATION NUMBER: US/09290049  
; SEQUENCE 16, Application US/09290049  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CRIES  
; FILE REFERENCE: FDC98-01p2A  
; CURRENT APPLICATION NUMBER: US/09/290,049  
; PRIOR FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081,550  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/115,142  
; EARLIER FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 16  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. mutans  
US-09-290-049-16

Query Match 81.8%; Score 90; DB 16; Length 22;  
Best Local Similarity 90.0%; Pred. No. 3e-08; 1; Indels 0; Gaps 0;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NYVFIRAHDSVQVTRIAKII 22  
||:||||||||| |||||  
Db 3 NYIFIRAHDSVQVTRIAKII 22

RESULT 3  
US-09-649-885-2  
; Sequence 2, Application US/09649885  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; TITLE OF INVENTION: Latexes in Paper Manufacture  
; FILE REFERENCE: 0358D2  
; CURRENT APPLICATION NUMBER: US/09/649,885  
; CURRENT FILING DATE: 2000-08-28  
; PRIOR APPLICATION NUMBER: US 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-649-885-2

Query Match 81.8%; Score 90; DB 20; Length 1430;  
Best Local Similarity 90.0%; Pred. No. 5.2e-06;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NYVFIRAHDSVQVTRIAKII 22  
||:||||||||| |||||  
Db 576 NYIFIRAHDSVQVTRIAKII 595

RESULT 4  
US-09-740-274-6  
; Sequence 6, Application US/09740274  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD

; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 6  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-740-274-6

Query Match 81.8%; Score 90; DB 21; Length 1430;  
Best Local Similarity 90.0%; Pred. No. 5.2e-06;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NYVFIRAHDSVQVTRIAKII 22  
||:||||||||| |||||  
Db 576 NYIFIRAHDSVQVTRIAKII 595

RESULT 5  
US-09-557-848-2  
; Sequence 2, Application US/09557848  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; TITLE OF INVENTION: Latexes in Paper Manufacture  
; FILE REFERENCE: 0356D2  
; CURRENT APPLICATION NUMBER: US/09/557,848  
; CURRENT FILING DATE: 2000-04-26  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-557-848-2

Query Match 72.7%; Score 80; DB 19; Length 1475;  
Best Local Similarity 77.3%; Pred. No. 0.00031;  
Matches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPNYVFIRAHDSVQVTRIAKII 22  
||:||||||||| |||||  
Db 552 VPSYFIRAHDSVQVTRIAKII 573

RESULT 6  
US-09-740-274-2  
; Sequence 2, Application US/09740274  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD

```

; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
; US-740-274-2

```

```

Query Match          72.7%; Score 80; DB 21; Length 1475;
Best Local Similarity 77.3%; Pred. No. 0.00031;
Matches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 1 VPNYVFIRAHDSVQTRIADII 22
   ||| ||||| ||||| ||| ||
Db 552 VPSYSFIRAHDSVQDLIRNII 573

```

```

RESULT 7
US-09-290-049-15
; Sequence 15, Application US/092900049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: FDC98-01p2A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 22
; TYPE: PRT
; ORGANISM: S. mutans
; US-09-290-049-15

```

```

Query Match          69.1%; Score 76; DB 16; Length 22;
Best Local Similarity 72.7%; Pred. No. 8.8e-06;
Matches 16; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 VPNYVFIRAHDSVQTRIADII 22
   ||| ||||| ||||| ||| ||
Db 1 VPSYSFIRAHDSVQDLIRNII 22

```

```

RESULT 8
US-09-740-274-4
; Sequence 4, Application US/09740274
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper

```

```

; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
; US-09-740-274-4

```

```

Query Match          69.1%; Score 76; DB 21; Length 1375;
Best Local Similarity 72.7%; Pred. No. 0.0015;
Matches 16; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 VPNYVFIRAHDSVQTRIADII 22
   ||| ||||| ||||| ||| ||
Db 578 VPSYSFIRAHDSVQDLIRNII 599

```

```

RESULT 9
US-09-290-049-17
; Sequence 17, Application US/092900049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: FDC98-01p2A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 22
; TYPE: PRT
; ORGANISM: S. downei
; US-09-290-049-17

```

```

Query Match          63.6%; Score 70; DB 16; Length 22;
Best Local Similarity 68.2%; Pred. No. 0.0001;
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

```

```

QY 1 VPNYVFIRAHDSVQTRIADII 22
   ||| ||||| ||||| ||| ||
Db 1 VPSYSFIRAHDSVQDLIRNII 22

```

```

RESULT 10
US-09-290-049-19
; Sequence 19, Application US/092900049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.

```

; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CARRIES  
; FILE REFERENCE: FDC98-01p2A  
; CURRENT APPLICATION NUMBER: US/09/290.049  
; CURRENT FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081.550  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/115.142  
; EARLIER FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 19  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. sobrinus  
US-09-290-049-19

Query Match 62.7%; Score 69; DB 16; Length 22;  
Best Local Similarity 68.2%; Pred. No. 0.00015;  
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 VPNYVFIRAHDSVQTRIARI 22  
Db 1 VPSYSFIRAHDSVQDIIARDII 22

RESULT 11  
US-09-499-203-2  
; Sequence 2, Application US/09499203  
; GENERAL INFORMATION:  
; APPLICANT: KOSSMANN, Jens  
; APPLICANT: WELSH, Thomas  
; APPLICANT: QUANZ, Martin  
; APPLICANT: KNUTH, Karola  
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase  
; FILE REFERENCE: 147-196P  
; CURRENT APPLICATION NUMBER: US/09/499.203  
; CURRENT FILING DATE: 2000-02-08  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2057  
; TYPE: PRT  
; ORGANISM: Leuconostoc mesenteroides  
9-499-203-2

Query Match 57.3%; Score 63; DB 18; Length 2057;  
Best Local Similarity 55.0%; Pred. No. 0.47;  
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 VPNYVFIRAHDSVQTRIARI 20  
Db 757 IPNYSFVRAHDYDAQDPIRK 776

RESULT 12  
US-09-290-049-2  
; Sequence 2, Application US/09290049  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CARRIES  
; FILE REFERENCE: FDC98-01p2A  
; CURRENT APPLICATION NUMBER: US/09/290.049  
; CURRENT FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081.550  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/115.142  
; EARLIER FILING DATE: 1999-01-08

; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: HDS peptide  
US-09-290-049-2

Query Match 56.8%; Score 62.5; DB 16; Length 20;  
Best Local Similarity 75.0%; Pred. No. 0.0019;  
Matches 15; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Qy 1 VPNYVFIR-AHDSVQTRIARI 19  
Db 1 VPSYSFIRAHDSVQDLIA\*20

RESULT 13  
US-60-161-932-1239  
; Sequence 1239, Application US/60161932  
; GENERAL INFORMATION:  
; APPLICANT: Venter, J. Craig  
; TITLE OF INVENTION: The Drosophila Genome, Primary Nucleic  
; TITLE OF INVENTION: Acid Sequences, Systems Containin the Nucleic Acid Sequences  
; TITLE OF INVENTION: and Uses Thereof  
; FILE REFERENCE: CL000122  
; CURRENT APPLICATION NUMBER: US/60/161.932  
; CURRENT FILING DATE: 1999-10-28  
; NUMBER OF SEQ ID NOS: 2626  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 1239  
; LENGTH: 336  
; TYPE: PRT  
; ORGANISM: Drosophila  
US-60-161-932-1239

Query Match 42.7%; Score 47; DB 24; Length 336;  
Best Local Similarity 42.1%; Pred. No. 33;  
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 3 NYVEIRAHDSVQTRIARI 21  
Db 185 NVAFTRAHNSDQPKLIQM 203

RESULT 14  
US-60-167-217-3108  
; Sequence 3108, Application US/60167217  
; GENERAL INFORMATION:  
; APPLICANT: Li, Peter W. D.  
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: CL000152  
; CURRENT APPLICATION NUMBER: US/60/167.217  
; CURRENT FILING DATE: 1999-11-24  
; NUMBER OF SEQ ID NOS: 23195  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 3108  
; LENGTH: 336  
; TYPE: PRT  
; ORGANISM: Drosophila  
US-60-167-217-3108

Query Match 42.7%; Score 47; DB 24; Length 336;  
Best Local Similarity 42.1%; Pred. No. 33;  
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 3 NYVFIRAHDSVQTRIARI 21  
| | | | | : | | : :  
Db 185 NVAFTRAHNSDQQTCLIQM 203

RESULT 15  
US-60-173-464-2558  
; Sequence 2558, Application US/60173464  
; GENERAL INFORMATION:  
; APPLICANT: Li, Peter W.D.  
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES  
; FILE REFERENCE: CL000173  
; CURRENT APPLICATION NUMBER: US/60/173,464  
; CURRENT FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 30269  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2558  
; LENGTH: 336  
; TYPE: PRT  
; ORGANISM: Drosophila  
; 0-173-464-2558

Query Match 42.7%; Score 47; DB 24; Length 336;  
Best Local Similarity 42.1%; Pred. No. 33;  
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 3 NYVFIRAHDSVQTRIARI 21  
| | | | | : | | : :  
Db 185 NVAFTRAHNSDQQTCLIQM 203

Search completed: March 27, 2002, 14:20:30  
Job time: 1577 sec





Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	110	100.0	22	6	US-09-562-328-31		Sequence 31, Appl
2	110	100.0	22	6	US-09-290-049A-18		Sequence 18, Appl
3	90	81.8	22	6	US-09-562-328-29		Sequence 29, Appl
4	90	81.8	22	6	US-09-290-049A-16		Sequence 16, Appl
5	90	81.8	545	6	US-09-604-957-4		Sequence 4, Appl
6	89	80.9	523	6	US-09-604-957-5		Sequence 5, Appl
7	76	69.1	22	6	US-09-562-328-28		Sequence 28, Appl
8	76	69.1	22	6	US-09-290-049A-15		Sequence 15, Appl
9	73	66.4	19	6	US-09-562-328-27		Sequence 27, Appl
10	73	66.4	19	6	US-09-290-049A-2		Sequence 2, Appl
11	73	66.4	23	6	US-09-562-328-44		Sequence 44, Appl
12	70	63.6	22	6	US-09-562-328-30		Sequence 30, Appl
13	70	63.6	22	6	US-09-290-049A-17		Sequence 17, Appl
14	69	62.7	22	6	US-09-562-328-32		Sequence 32, Appl
15	69	62.7	22	6	US-09-290-049A-19		Sequence 19, Appl
16	65	59.1	584	6	US-09-604-957-6		Sequence 6, Appl
17	63	57.3	535	6	US-09-604-957-7		Sequence 7, Appl
18	63	57.3	1278	6	US-09-604-957-3		Sequence 3, Appl
19	62.5	56.8	20	6	US-09-562-328-26		Sequence 26, Appl
20	47	42.4	336	6	US-09-614-150-3054		Sequence 3054, Ap
21	45.5	41.4	866	6	US-09-614-150-21291		Sequence 21291, A
22	45	40.9	639	6	US-09-920-954-4		Sequence 4, Appl
23	44	40.0	575	6	US-09-708-427-10938		Sequence 10938, A
24	44	40.0	577	6	US-09-708-427-10937		Sequence 10937, A
25	44	40.0	583	6	US-09-708-427-10936		Sequence 10936, A

; SEQ ID NO 18  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. downei  
US-09-290-049A-18

Query Match 100.0%; Score 110; DB 6; Length 22;  
Best Local Similarity 100.0%; Pred. No. 5.6e-12;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VNVVFIHDSVQVTRIAKII 22  
||:|||||||||  
Db 1 VNVVFIHDSVQVTRIAKII 22

RESULT 3  
US-09-562-328-29  
; Sequence 29, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 29  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-29

Query Match 81.8%; Score 90; DB 6; Length 22;  
Best Local Similarity 90.0%; Pred. No. 1.3e-08;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 NVVFIHDSVQVTRIAKII 22  
||:|||||||||  
Db 3 NYIFIRHDSVQVTRIAKII 22

RESULT 4  
US-09-290-049A-16  
; Sequence 16, Application US/09290049A  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 16  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. mutans  
US-09-290-049A-16

Query Match 81.8%; Score 90; DB 6; Length 22;  
Best Local Similarity 90.0%; Pred. No. 1.3e-08;

Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 3 NYVFIHDSVQVTRIAKII 22  
||:|||||||||  
Db 3 NYIFIRHDSVQVTRIAKII 22

RESULT 5  
US-09-604-957-4  
; Sequence 4, Application US/09604957  
; GENERAL INFORMATION:  
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
; APPLICANT: DIJKHUIZEN, LUBBERT  
; APPLICANT: RAHAOUI, HAKIM  
; APPLICANT: LEER, ROBERT-JAN  
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
; FILE REFERENCE: BO 43388  
; CURRENT APPLICATION NUMBER: US/09/604,957  
; CURRENT FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 00201871.1  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 545  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-604-957-4

Query Match 81.8%; Score 90; DB 6; Length 545;  
Best Local Similarity 90.0%; Pred. No. 6.9e-07;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 NYVFIHDSVQVTRIAKII 22  
||:|||||||||  
Db 156 NYIFIRHDSVQVTRIAKII 175

RESULT 6  
US-09-604-957-5  
; Sequence 5, Application US/09604957  
; GENERAL INFORMATION:  
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
; APPLICANT: DIJKHUIZEN, LUBBERT  
; APPLICANT: RAHAOUI, HAKIM  
; APPLICANT: LEER, ROBERT-JAN  
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
; FILE REFERENCE: BO 43388  
; CURRENT APPLICATION NUMBER: US/09/604,957  
; CURRENT FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 00201871.1  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 523  
; TYPE: PRT  
; ORGANISM: Leuconostoc mesenteroides  
US-09-604-957-5

Query Match 80.9%; Score 89; DB 6; Length 523;  
Best Local Similarity 72.7%; Pred. No. 9.6e-07;  
Matches 16; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 VNVVFIHDSVQVTRIAKII 22  
||:|||||||||  
Db 146 IPNYSFVRHDSVQVTRIAQIV 167

RESULT 7  
US-09-562-328-28

; Sequence 28, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 28  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-28

Query Match 69.1%; Score 76; DB 6; Length 22;  
Best Local Similarity 72.7%; Pred. No. 2.8e-06;  
Matches 16; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
QY 1 VPVYVFIRAHDSVQTRIAKII 22  
Db 1 VPSYSFIRAHDSVQDLIRNII 22

RESULT 8  
US-09-290-049A-15  
; Sequence 15, Application US/09290049A  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CARIES  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. mutans  
9-290-049A-15

Query Match 69.1%; Score 76; DB 6; Length 22;  
Best Local Similarity 72.7%; Pred. No. 2.8e-06;  
Matches 16; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
QY 1 VPVYVFIRAHDSVQTRIAKII 22  
Db 1 VPSYSFIRAHDSVQDLIRNII 22

RESULT 9  
US-09-562-328-27  
; Sequence 27, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01

; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 27  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-27

Query Match 66.4%; Score 73; DB 6; Length 19;  
Best Local Similarity 78.9%; Pred. No. 7.4e-06;  
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 VPVYVFIRAHDSVQTRIA 19  
Db 1 VPSYSFIRAHDSVQDLIA 19

RESULT 10  
US-09-290-049A-2  
; Sequence 2, Application US/09290049A  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CARIES  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: HDS peptide  
US-09-290-049A-2

Query Match 66.4%; Score 73; DB 6; Length 19;  
Best Local Similarity 78.9%; Pred. No. 7.4e-06;  
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 VPVYVFIRAHDSVQTRIA 19  
Db 1 VPSYSFIRAHDSVQDLIA 19

RESULT 11  
US-09-562-328-44  
; Sequence 44, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 44  
; LENGTH: 23  
; TYPE: PRT

; ORGANISM: Streptococcus sp.  
US-09-562-328-44

Query Match 66.4%; Score 73; DB 6; Length 23;  
Best Local Similarity 78.9%; Pred. No. 9.3e-06;  
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPNNVFIRAHDSVQTRIA 19  
Db 5 VPSYFIRAHDSVQDLRIA 23

RESULT 12  
US-09-562-328-30  
; Sequence 30, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 30  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-30

Query Match 63.6%; Score 70; DB 6; Length 22;  
Best Local Similarity 68.2%; Pred. No. 2.8e-05;  
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 VPNNVFIRAHDSVQTRIAKII 22  
Db 1 VPSYFIRAHDSVQDLRIIDII 22

RESULT 13  
US-09-290-049A-17  
; Sequence 17, Application US/09290049A  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. downei  
US-09-290-049A-17

Query Match 63.6%; Score 70; DB 6; Length 22;  
Best Local Similarity 68.2%; Pred. No. 2.8e-05;  
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 VPNNVFIRAHDSVQTRIAKII 22

Db 1 VPSYFIRAHDSVQDLRIIDII 22

RESULT 14  
US-09-562-328-32  
; Sequence 32, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 32  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-32

Query Match 62.7%; Score 69; DB 6; Length 22;  
Best Local Similarity 58.2%; Pred. No. 4.1e-05;  
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 VPNNVFIRAHDSVQTRIAKII 22  
Db 1 VPSYFIRAHDSVQDLRIIDII 22

RESULT 15  
US-09-290-049A-19  
; Sequence 19, Application US/09290049A  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 19  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. sobrinus  
US-09-290-049A-19

Query Match 62.7%; Score 69; DB 6; Length 22;  
Best Local Similarity 58.2%; Pred. No. 4.1e-05;  
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 VPNNVFIRAHDSVQTRIAKII 22  
Db 1 VPSYFIRAHDSVQDLRIIDII 22

Search completed: March 27, 2002, 14:22:49  
Job time: 1696 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:26:08 ; Search time 188.53 Seconds  
(without alignments)  
17.069 Million cell updates/sec

Title: US-09-290-049a-16  
Perfect score: 107  
Sequence: 1 MANYIFIRAHSEVQTVIKII 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
1 number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phage:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	100	93.5	1575	2 Q9LCH3	Q9Lch3 streptococc
2	95	88.8	1577	2 Q54178	Q54178 streptococc
3	90	84.1	1338	2 Q9WXJ4	Q9wxj4 streptococc
4	89	83.2	1577	2 Q5265	Q5265 streptococc
5	88	82.2	1449	2 Q68542	Q68542 streptococc
6	88	82.2	1449	2 Q5264	Q5264 streptococc
7	87	81.3	1016	2 Q9LCJ7	Q9lcj7 leuconostoc
8	87	81.3	1477	2 Q5L466	Q5l466 leuconostoc
9	87	81.3	1508	2 Q52224	Q52224 leuconostoc
10	87	81.3	1508	2 Q9EZH5	Q9ezh5 leuconostoc
11	87	81.3	1512	2 Q9WXJ5	Q9wxj5 streptococc
12	87	81.3	1599	2 Q00599	Q00599 streptococc
13	86	80.4	1527	2 Q9ZAR4	Q9zar4 leuconostoc
14	85	79.4	1290	2 Q48756	Q48756 leuconostoc
15	78	72.9	1518	2 Q00600	Q00600 streptococc
16	68	63.6	1455	2 Q69388	Q69388 streptococc
17	67	62.6	1390	2 Q69385	Q69385 streptococc
18	67	62.6	1455	2 Q69382	Q69382 streptococc
19	67	62.6	1455	2 Q69391	Q69391 streptococc

20	67	62.6	1455	2 Q69397	Q69397 streptococc
21	64	59.8	1590	2 Q55263	Q55263 streptococc
22	64	59.8	1590	2 Q5983	Q5983 streptococc
23	53	49.5	2057	2 Q9RE05	Q9re05 leuconostoc
24	47	43.9	474	5 Q9VCH8	Q9vch8 drosophila
25	47	43.9	1197	5 Q9BI21	Q9bi21 drosophila
26	46.5	43.5	866	5 Q9V8T7	Q9v8t7 drosophila
27	45	42.1	267	1 Q9HRT8	Q9hrt8 halobacteri
28	45	42.1	437	2 Q99YP6	Q99yp6 streptococc
29	45	42.1	665	5 Q9V4L1	Q9v4l1 drosophila
30	44	41.1	132	2 Q9KPH6	Q9kph6 vibrio chol
31	44	41.1	341	2 Q9L7T7	Q9l7t7 rhodospiril
32	44	41.1	476	10 Q9LUR9	Q9lur9 arabidopsis
33	44	41.1	583	10 Q49505	Q49505 arabidopsis
34	44	41.1	694	2 Q9L5M3	Q9l5m3 salmonella
35	44	41.1	1213	3 Q59801	Q59801 schizosacch
36	43	40.2	261	1 P94945	P94945 methanopyru
37	43	40.2	288	5 Q9NAM7	Q9nam7 caenorhabdi
38	43	40.2	933	2 Q51486	Q51486 borrelia bu
39	41.5	38.8	271	2 Q33982	Q33982 borrelia bu
40	41.5	38.8	354	2 Q50835	Q50835 borrelia bu
41	41	38.3	76	12 Q96894	Q96894 human herpe
42	41	38.3	111	10 Q9SPK9	Q9spk9 lactuca sat
43	41	38.3	140	2 Q07750	Q07750 mycobacteri
44	41	38.3	196	10 Q9FUL2	Q9ful2 prunus aviu
45	41	38.3	348	10 Q9ZR34	Q9zr34 triticum ae

ALIGNMENTS

RESULT 1

ID Q9LCH3 PRELIMINARY; PRT; 1575 AA.

AC Q9LCH3;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE GLUCOSYLTRANSFERASE.

GN GTFR.

OS Streptococcus oralis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1303;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC10557;

RA MEDLINE=20231779; PubMed=10768934;

RA Fujiwara T., Hoshino T., Ooshima T., Sobue S., Hamada S.;

RT "Purification, characterization, and molecular analysis of the gene

RT encoding glucosyltransferase from Streptococcus oralis.";

RL Infect. Immun. 68:2475-2483(2000).

DR EMBL; AB025228; BAA95201.1; -.

DR InterPro; IPR002479; CW\_binding.

DR	InterPro; IPR003318; Glyco\_hydro\_70.
DR	Pfam; PF01473; CW\_binding\_1; 17.
DR	Pfam; PF02324; Glyco\_hydro\_70; 1.
DR	Transferrase.
DR	SEQUENCE 1575 AA; 176792 MW; 772A26E4D7C2E543 CRC64;

Query Match 93.5%; Score 100; DB 2; Length 1575;  
Best Local Similarity 90.9%; Pred.No. 7.8e-08;  
Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 MANYIFIRAHSEVQTVIKII 22  
|||||:|||||  
Db 617 MANYIFVRAHSEVQTVIADII 638

RESULT 2

Q54178 PRELIMINARY; PRT; 1577 AA.

ID Q54178

AC Q54178; Q54247;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE.  
GN GTF.  
OS Streptococcus gordonii Challis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=29390;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CHALLIS;  
RX MEDLINE=96157084; PubMed=8586195;  
RA Vickerman M.M., Sulavik M.C., Clewell D.B.:  
RT "Molecular analysis of Streptococcus gordonii glucosyltransferase  
phase variants.";  
RL Dev. Biol. Stand. 85:309-314(1995).  
[2]  
RC SEQUENCE OF 1-96 FROM N.A.  
RC STRAIN=CHALLIS;  
RX MEDLINE=92276337; PubMed=1534326;  
RA Sulavik M.C., Tardif G., Clewell D.B.:  
RT "Identification of a gene, rgg, which regulates expression of  
glucosyltransferase and influences the spp phenotype of Streptococcus  
gordonii Challis.";  
RL J. Bacteriol. 174:3577-3586(1992).  
DR EMBL; U12643; AAC43483.1; -  
DR EMBL; M89776; AAA26969.1; -  
DR InterPro; IPR002479; CW\_binding.  
DR Pfam; PF01473; Glyco\_hydro\_70.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase  
SQ SEQUENCE 1577 AA; 177805 MW; 5AE0328DC5E08D18 CRC64;

Query Match 88.8%; Score 95; DB 2; Length 1577;  
Best Local Similarity 90.5%; Pred. No. 5.4e-07;  
Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 ANYIFRAHDSVQTVIAKII 22  
|||||:|||||:|||||:|  
Db 620 ANYIFRAHDSVQTVIADII 640

RESULT 3  
J4  
AC Q9WXJ4 PRELIMINARY; PRT; 1338 AA.  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE GTF-S.  
GN GTF.  
OS Streptococcus criceti.  
OG plasmid pAM1.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1333;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HS-6;  
RA Inoue M., Fukui K., Miyagi A.:  
RT "S.cricetus glucosyltransferase(gtfs and gtfT) genes.";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB026123; BAA77236.1; -  
DR HSSP; P06278; 1VJS.  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 10.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Plasmid.

SQ SEQUENCE 1338 AA; 148558 MW; 0A90C8E10E15D99B CRC64;

Query Match 84.1%; Score 90; DB 2; Length 1338;  
Best Local Similarity 90.0%; Pred. No. 3.1e-06;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 NYIFRAHDSVQTVIAKII 22  
|||||:|||||:|||||:|  
Db 511 NYIFRAHDSVQTVIAKII 530

RESULT 4  
Q55265 PRELIMINARY; PRT; 1577 AA.  
ID Q55265;  
AC Q55265;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE PRECURSOR.  
GN GTF.  
OS Streptococcus salivarius.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1304;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95122197; PubMed=7822030;  
RA Simpson C.L., Giffard P.M., Jacques N.A.:  
RT "Streptococcus salivarius ATCC 25975 possesses at least two genes  
coding for primer-independent glucosyltransferases.";  
RL Infect. Immun. 63:609-621(1995).  
DR EMBL; L35928; AAC41413.1; -  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 11.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
FT SIGNAL; 1 38 POTENTIAL.  
FT CHAIN 39 1577 GLUCOSYLTRANSFERASE.  
SQ SEQUENCE 1577 AA; 175290 MW; 3EFB898A7D3A7BF3 CRC64;

Query Match 83.2%; Score 89; DB 2; Length 1577;  
Best Local Similarity 81.0%; Pred. No. 5.4e-06;  
Matches 17; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 ANYIFRAHDSVQTVIAKII 22  
|||||:|||||:|||||:|  
Db 660 ANYIFRAHDSVQAVLANII 680

RESULT 5  
O68542 PRELIMINARY; PRT; 1449 AA.  
ID O68542;  
AC O68542;  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE N (FRAGMENT).  
GN GTF.  
OS Streptococcus salivarius.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1304;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=V1477;  
RA Jaffe R.I.;  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF049609; AAC05156.1; -  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.



DR Pfam; PF01473; CW\_binding.1; 8.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase.  
 FT NON\_TER 1449 1449  
 SQ SEQUENCE 1449 AA; 159895 MW; 0700F6D748471BFB CRC64;

Query Match 82.2%; Score 88; DB 2; Length 1449;  
 Best Local Similarity 72.7%; Pred. No. 7.2e-06;  
 Matches 16; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 MANYFIRAHDSVQTVIAKII 22  
 ||||| :|||||:|:|:  
 Db 607 MANYAFVRAHDSVQSIIGQII 628

RESULT 6  
 Q55264 ID Q55264 PRELIMINARY; PRT; 1449 AA.  
 AC Q55264;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE PRECURSOR.  
 GN GFL.  
 OS Streptococcus salivarius.  
 OS Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OC NCBI\_TaxID=1304;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95122197; PubMed=7822030;  
 RA Simpson C.L., Giffard P.M., Jacques N.A.;

RT "Streptococcus salivarius ATCC 25975 possesses at least two genes  
 coding for primer-independent glucosyltransferases.";  
 RL Infect. Immun. 63:609-621(1995).  
 DR EMBL; L35495; AAC41412.1; -

DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding.1; 8.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Signal; transferase.  
 FT SIGNAL 1 35 POTENTIAL.  
 FT CHAIN 36 1449 GLUCOSYLTRANSFERASE.  
 SQ SEQUENCE 1449 AA; 159984 MW; DD62F07306E86A46 CRC64;

Query Match 82.2%; Score 88; DB 2; Length 1449;  
 Best Local Similarity 72.7%; Pred. No. 7.2e-06;  
 Matches 16; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 MANYFIRAHDSVQTVIAKII 22  
 ||||| :|||||:|:|:  
 Db 607 MANYAFVRAHDSVQSIIGQII 628

RESULT 7  
 Q9LCJ7 ID Q9LCJ7 PRELIMINARY; PRT; 1016 AA.  
 AC Q9LCJ7;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE DEXTRANSUCRASE.  
 GN DSRT.  
 OS Leuconostoc mesenteroides.  
 OS Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC Leuconostoc.  
 OC NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL B-512F;  
 RX MEDLINE=20169623; PubMed=10705445;

RA Funane K., Mizuno K., Takahara H., Kobayashi M.;  
 RT "Gene encoding a dextranucrase-like protein in Leuconostoc  
 mesenteroides NRRL B-512F.";  
 RL Biosci. Biotechnol. Biochem. 64:29-38(2000).  
 DR EMBL; AB020020; BAA90527.1; -  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 SQ SEQUENCE 1016 AA; 110343 MW; 8896EFDE13CCCB47 CRC64;

Query Match 81.3%; Score 87; DB 2; Length 1016;  
 Best Local Similarity 85.0%; Pred. No. 7.2e-06;  
 Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 NYIFRAHDSVQTVIAKII 22  
 ||| :|||||:|:|:  
 Db 626 NYSFVRAHDSVQTVIAEII 645

RESULT 8  
 Q9L466 ID Q9L466 PRELIMINARY; PRT; 1477 AA.  
 AC Q9L466;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE DEXTRANSUCRASE (EC 2.4.1.5).  
 GN DSRC.

OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC Leuconostoc.  
 OC NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=NRRL B-1355;  
 RA Arguello-Morales M.A., Remaud-Simeon M., Pizzut S., Sarcabal P.,  
 RA Willemot R.M., Monsan P.;  
 RT "Sequence analysis of the gene encoding alternansucrase, a sucrose  
 glucosyltransferase from Leuconostoc mesenteroides NRRL B-1355.";  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ250172; CAB/6565.1; -

DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding.1; 14.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase; Glycosyltransferase.  
 SQ SEQUENCE 1477 AA; 164886 MW; E6F5710DEDFCB831 CRC64;

Query Match 81.3%; Score 87; DB 2; Length 1477;  
 Best Local Similarity 85.0%; Pred. No. 1.1e-05;  
 Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 NYIFRAHDSVQTVIAKII 22  
 ||| :|||||:|:|:  
 Db 605 NYSFVRAHDSVQTVIAQII 624

RESULT 9  
 O52224 ID O52224 PRELIMINARY; PRT; 1508 AA.  
 AC O52224;

DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE 6-  
 DE GLUCOSYLTRANSFERASE).  
 GN DSRR.

OS Leuconostoc mesenteroides.  
 OS Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC Leuconostoc.  
 OC NCBI\_TaxID=1245;  
 RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-NRRL B-1299;  
RA Monchois V., Renaud-Simeon M., Monsan P., Willemot R.M.;  
RL FEMS Microbiol. Lett. 0:0-0(1998).  
CC -!- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) = D-FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).  
DR EMBL; AF030129; AAB95453.1; -.  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 14.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase; Glycosyltransferase.  
SQ SEQUENCE 1508 AA; 168511 MW; E70CECB57A70DIF0 CRC64;

Query Match 81.3%; Score 87; DB 2; Length 1508;  
Best Local Similarity 85.0%; Pred. No. 1.1e-05;  
Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

3 NYIFIRAHDSVQTVIAKII 22  
II I:|||||:|||||:II  
Db 636 NYSFVRAHDSVQTVIAKII 655

RESULT 10  
Q9EZH5 PRELIMINARY; PRT; 1508 AA.  
AC Q9EZH5;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE DEXTRANSUCRASE DSRB742.  
GN DSRB742.  
OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OC Leuconostoc.  
OX NCBI\_TaxID=1245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-B-742CB;  
RA Kim H.-S., Kim D., Ryu H.-J., Robyt J.F.;  
RL "Leuconostoc mesenteroides B-742CB, a dextransucrase gene.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF294469; AAG38021.1; -.  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 14.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
SQ SEQUENCE 1508 AA; 168542 MW; E2FCFA0F87AE4F3A CRC64;

Query Match 81.3%; Score 87; DB 2; Length 1508;  
Best Local Similarity 85.0%; Pred. No. 1.1e-05;  
Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

3 NYIFIRAHDSVQTVIAKII 22  
II I:|||||:|||||:II  
Db 636 NYSFVRAHDSVQTVIAKII 655

RESULT 11  
Q9WXJ5 PRELIMINARY; PRT; 1512 AA.  
AC Q9WXJ5;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE GTF-S.  
GN GTF.  
OS Streptococcus criceti.  
OG Plasmid pAMI.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.

OX NCBI\_TaxID=1333;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-HS-6;  
RA Inoue M., Fukui K., Miyagi A.;  
RT "S.cricetus glucosyltransferase(gtfs and gtfT) genes.";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB026123; BAA77237.1; -.  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 14.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Plasmid.  
SQ SEQUENCE 1512 AA; 167145 MW; 4C03D9CBC601FC14 CRC64;

Query Match 81.3%; Score 87; DB 2; Length 1512;  
Best Local Similarity 80.0%; Pred. No. 1.1e-05;  
Matches 16; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

3 NYIFIRAHDSVQTVIAKII 22  
:I:I:|||||:|||||:II  
Db 561 SYFVRAHDSVQTVIAKII 580

RESULT 12  
Q00599 PRELIMINARY; PRT; 1599 AA.  
ID Q00599  
AC Q00599;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE S PRECURSOR (EC 2.4.1.5) (GTF) (DEXTRANSUCRASE)  
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
GN GTFK.  
OS Streptococcus salivarius.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1304;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 25975;  
RX MEDLINE=93381463; PubMed=8371114;  
RA Giffard P.M., Allen D.M., Milward C.P., Simpson C.L., Jacques N.A.;  
RT "Sequence of the gtfK gene of Streptococcus salivarius ATCC 25975 and evolution of the gtf genes of oral streptococci.";  
RL J. Gen. Microbiol. 137:2577-2593(1991).  
CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS. THAT ARE THOUGHT TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -!- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) = D-FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).  
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.  
CC -!- DISEASE: DENTAL CARIES.  
CC -!- SIMILARITY: TO REGIONS OF BARLEY AND BACILLUS AMYLOLIQUEFACIENS ALPHA AMYLASES AND RABBIT GLYCOGEN PHOSPHORYLASE.  
DR EMBL; Z11872; CAA77898.1; -.  
DR EMBL; Z11873; CAA77901.1; -.  
DR EMBL; M64111; AAA26897.1; -.  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 14.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.

RA Monchois V., Willemot R.M., Renaud-Simeon M., Croux C., Monsan P.;  
RT "Cloning and sequencing of a gene coding for a novel dextranucrase  
RT from *Leuconostoc mesenteroides* NRRL B-1299 synthesizing only alpha (1-  
RT 6) and alpha (1-3) linkages.";

Search completed: March 27, 2002, 14:26:09  
Job time: 1681 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:57:59 ; Search time 198.55 Seconds  
(without alignments)  
8.208 Million cell updates/sec

Title: US-09-290-049a-17  
Perfect score: 110  
Sequence: 1 VPSYFARAHDSVQDLRDII 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues  
1 number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_1101.\*  
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	108	98.2	1592	14	AA1980
2	68	61.8	2057	21	AA1981
3	62	56.4	1577	17	AA1982
4	45	40.9	2625	19	AA1983
5	45	40.9	2627	19	AA1984
6	44	40.0	519	20	AA1985
7	44	40.0	743	22	AA1986
8	44	40.0	743	22	AA1987
9	43.5	39.5	484	18	AA1988
10	43	39.1	164	21	AA1989
11	43	39.1	166	22	AA1990

Zea mays protein f  
Zea mays protein f  
Interferon induced  
Interferon induced  
Interferon induced  
Interferon induced  
Zea mays protein f  
Human atrial natri  
Atrial natriuretic  
Trypanosoma cruzi  
Human polyprotein  
Arabidopsis thalia  
Arabidopsis thalia  
Corn cyclin-depend  
Zea mays protein f  
Zea mays protein f  
RstA protein of CT  
Platelet aggregati  
B. burgdorferi ant  
B. burgdorferi ant  
Urea amidolyase.  
Zea mays protein f  
Bacillus alkaline  
Alpha-ketoglutarat  
Succinate dehydrog  
H. pylori GHPO 576  
H. pylori ORF 029e  
Neuronal nitric ox  
Human colon cancer  
Growth hormone act  
A human proliferat  
H. pylori ORF 03ge  
H. pylori outer me

ALIGNMENTS

RESULT 1  
AA1980  
ID AA1980 standard; Protein; 1592 AA.  
AC AA1980;  
XX AA1980;  
XX AA1980;  
DT 28-JUN-1993 (first entry)  
XX Glucosyltransferase I.  
DE Glucosyltransferase I.  
XX GT-1; Streptococcus; dental; caries.  
XX Streptococcus sobrinus.  
XX JP05023188-A.  
XX 02-FEB-1993.  
XX 25-JUL-1991; 91JP-0186592.  
XX 25-JUL-1991; 91JP-0186592.  
XX (FUKU/) FUKUI I.  
XX (KATO/) KATO K.  
XX WPI; 1993-079449/10.  
XX N-PSDB; AAQ37760.  
XX DNA sequence glucosyl:transferase-I - comprises Streptococcus  
XX sobrinus DNA sequence with at least one nucleotide added or  
XX deleted  
XX Claim 13; Page 15; 29pp; Japanese.

CC The DNA sequence from Streptococcus sobrinus strain 6715 encodes  
 CC glucosyltransferase-I (and mutants). The DNA was obtd. by treating  
 CC S. sobrinus 6715 with mutanolysin, extracting the chromosomal DNA,  
 CC partially digesting with Sau3AI and fractionating on agarose gel.  
 CC The 3-5 kbp fragment was ligated into pUC18 and E. coli JM109  
 CC transformed with it. A Gr-I expressing clone was isolated and  
 CC sequenced. The clone may be used in the development of a drug for  
 CC dental caries.  
 XX  
 SQ Sequence 1592 AA;

Query Match 98.2%; Score 108; DB 14; Length 1592;  
 Best Local Similarity 95.5%; Pred. No. 1.2e-09;  
 Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VPSYSFARAHSEVQDLIRDI 22  
 |||||  
 548 vpsysfarahdsevdqldir 569

RESULT 2  
 AAB10667  
 ID AAB10667 standard; Protein: 2057 AA.  
 XX  
 AC AAB10667;  
 XX  
 DT 19-JAN-2001 (first entry)  
 XX  
 DE L. mesenteroides alternan sucrose protein.  
 XX  
 KW Alternan sucrose; glucosyltransferase; fructose; cosmetic; foodstuff;  
 KW syrup.  
 XX  
 OS Leuconostoc mesenteroides.  
 XX  
 PN DE19905069-A1.  
 XX  
 PD 10-AUG-2000.  
 XX  
 PF 08-FEB-1999; 99DE-1005069.  
 XX  
 PR 08-FEB-1999; 99DE-1005069.  
 XX  
 PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH  
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

Kossmann J, Welsh T, Quanz M, Knuth K;  
 WPI; 2000-550294/51.  
 DR N-PSDB; AAA97904.  
 XX  
 PT New nucleic acid encoding recombinant Leuconostoc mesenteroides  
 PT alternan sucrose protein and methods of alternan and fructose  
 PT production -  
 XX  
 PS Claim 1a; Page 30-36; 64pp; German.

CC This invention describes a novel nucleic acid molecule (I) encoding an  
 CC alternan sucrose (E.C. 2.4.1.140 - an enzyme, that belongs to the  
 CC glucosyltransferase group) The recombinant, purified alternan sucrose  
 CC gene is useful for the fermentative production of alternan (a  
 CC carbohydrate) and/or fructose by secreting the enzyme into a  
 CC saccharose-containing culture medium. Alternatively, the enzyme is  
 CC contacted with a saccharose-containing solution. The alternan and/or  
 CC fructose is then isolated from the medium. Cosmetic products or  
 CC foodstuffs containing alternan can be produced. Recombinant production of  
 CC alternan sucrose is advantageous as it provides a cost effective means of  
 CC producing fructose for high fructose containing syrups, production of  
 CC which previously has been achieved by costly production from maize  
 CC starch. This sequence represents the Leuconostoc mesenteroides alternan  
 CC sucrose protein which is described in the method of the invention.  
 XX

SQ Sequence 2057 AA;  
 Query Match 61.8%; Score 68; DB 21; Length 2057;  
 Best Local Similarity 63.2%; Pred. No. 0.012;  
 Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 VPSYSFARAHSEVQDLIR 19  
 :|||: |||: |||:  
 Db 757 ipnysfrahdydqdpir 775

RESULT 3  
 AAR91047  
 ID AAR91047 standard; Protein: 1577 AA.  
 XX  
 AC AAR91047;  
 XX  
 DT 22-MAY-1996 (first entry)  
 XX  
 DE Alpha-D-glucosyltransferase.  
 XX  
 KW Alpha-D-glucosyltransferase; primer-independent; soluble glucan;  
 KW sucrose; transgenic plant; cloning; Escherichia coli;  
 KW phage lambda-Cl3; vector; plasmid pGSG501; plasmid pGSG502;  
 KW gene transfer; crop improvement; storage carbohydrate; pasture;  
 KW feedstuff; senescence; dextran; binder; food; pharmaceutical.  
 XX  
 OS Streptococcus salivarius strain ATCC 25975.  
 XX  
 PN WO9606173-A1.  
 XX  
 PD 29-FEB-1996.  
 XX  
 PF 24-AUG-1995; 95WO-AU00527.  
 XX  
 PR 24-AUG-1994; 94AU-0007643.  
 XX  
 PA (GIFF/) GIFFARD P M.  
 PA (JACQ/) JACQUES N A.  
 PA (SIMP/) SIMPSON C L.

PI Giffard PM, Jacques NA, Simpson CL;  
 XX  
 XX WPI; 1996-151376/15.  
 DR N-PSDB; AAT13139.  
 XX  
 PT Plants contg. new bacterial DNA encoding glucosyl transferase  
 PT activity - retain higher levels of stored carbohydrate(s) in a form  
 PT readily digestible by ruminants  
 XX  
 PS Claim 4; Page 16-20; 31pp; English.

CC The sequence represents an alpha-D-glucosyltransferase from  
 CC Streptococcus salivarius. The enzyme is primer-independent, and  
 CC produces soluble glucan from sucrose. A gene encoding the enzyme  
 CC may be cloned and expressed in Escherichia coli using a subclone  
 CC of phage lambda-Cl3, e.g. plasmid pGSG501 or plasmid pGSG502. The  
 CC DNA may also be expressed in a transgenic plant, to improve the  
 CC level of stored carbohydrate in a pasture plant, which normally  
 CC contains low levels, or to prevent degradation of stored carbohydrate  
 CC during plant senescence. Dextran may be isolated from the plant, for  
 CC use as a food binder or pharmaceutical additive. Primer independence  
 CC ensures that the enzyme will be functional in plants. The glucan is  
 CC poorly degraded in plants but easily degraded by bacteria in the rumen  
 CC of grazing livestock.  
 XX

SQ Sequence 1577 AA;

Query Match 56.4%; Score 62; DB 17; Length 1577;  
 Best Local Similarity 60.0%; Pred. No. 0.095;  
 Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 SYSFARAHSEVQDLIRDII 22  
:|||: :|||  
Db 661 nyifvrahdsqvavlanii 680

RESULT 4  
AAW55887  
ID AAW55887 standard; Protein; 2625 AA.  
XX  
AC AAW55887;  
XX 22-JUL-1998 (first entry)  
XX Human telomerase.  
XX  
XX Rat; telomerase; human; cancer; screening; inhibitor; elucidation;  
KW detection; probe; diagnosis; cell growth; ageing.  
XX  
XX Homo sapiens.  
XX WO9807838-A1.  
XX 26-FEB-1998.  
XX 21-AUG-1997; 97WO-JP02904.  
XX 17-FEB-1997; 97JP-0031807.  
XX 21-AUG-1996; 96JP-0219761.  
XX 31-JAN-1997; 97JP-0018878.  
XX (MITU) MITSUBISHI CHEM CORP.  
XX  
XX Fujino Y, Harada N, Ishikawa F, Nakamura H, Takahashi K;  
XX WPI; 1998-169149/15.  
XX N-PSDB; AAV25989.  
XX  
XX Telomerase protein of higher animals and humans and gene encoding it  
PT - for use in diagnosis of cancer, screening of telomerase inhibitors  
PT and elucidation of biological control mechanisms  
XX  
XX Claim 8; Page 74-93; 106pp; Japanese.  
XX  
XX The present sequence represents a human protein component of telomerase.  
XX The DNA or RNA encoding the telomerase probe for the detection of  
XX fragments can be used as a nucleotide probe for the detection of  
XX cancer cells and for diagnosis of cancer. Potential telomerase  
XX inhibitors can be screened by measuring their effect on the assay  
XX of the active form in cells or tissues. The polypeptide and DNA  
XX coding for it can be used in the elucidation of biological control  
XX mechanisms of, e.g. cell growth or ageing and of the mechanisms of  
XX cancer development.  
XX  
XX Sequence 2625 AA;

Query Match 40.9%; Score 45; DB 19; Length 2625;  
Best Local Similarity 47.1%; Pred. No. 1.5e+02;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 VPSYSFARAHSEVQDL 17  
:||||: :|||  
Db 205 mpsysisgeeevedl 221

RESULT 5  
AAW61347  
ID AAW61347 standard; Protein; 2627 AA.  
XX  
AC AAW61347;  
XX  
XX 12-OCT-1998 (first entry)

XX Human telomerase RNA interactive protein-1 (TRIP1).  
XX TRIP1; human; telomerase RNA interactive protein-1; cancer; AIDS;  
KW ageing; therapy.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 1155  
FT /note= "residue 1155 is given as Xaa in Fig 3  
FT (translates as Gln)"  
XX  
XX WO9821343-A1.  
XX  
XX 22-MAY-1998.  
XX  
XX 13-NOV-1997; 97WO-US21248.  
XX  
XX 16-OCT-1997; 97US-0951733.  
XX 15-NOV-1996; 96US-0871189.  
XX 11-JUN-1997; 97US-0873039.  
XX  
XX (AMGE-) AMGEN CANADA INC.  
XX (AMGE-) AMGEN INC.  
XX  
XX Harrington LA, Robinson MO;  
XX  
XX WPI; 1998-297946/26.  
XX N-PSDB; AAV27865.  
XX  
XX New nucleic acid encoding human telomerase protein-2 - used for  
XX regulating telomerase activity, e.g. for treating cancer or acquired  
XX immune deficiency syndrome  
XX  
XX Example 2; Fig 3; 150pp; English.  
XX  
XX This polypeptide comprises human telomerase RNA interactive  
XX protein-1 (TRIP1). Its amino acid sequence was deduced from  
XX cDNA clones (see AAV27865) obtained from a colon tumour cell line  
XX LIM1863 cDNA library. The invention relates to novel genes  
XX encoding components of the telomerase enzyme complex, including  
XX TRIP1 and telomerase protein 2 (TP2, see also AAW61349). These  
XX polypeptides may be useful as therapeutic agents in those cases  
XX where increasing TRIP1 activity or TP2 activity is desired, e.g.  
XX for treatment of HIV infection, AIDS and ageing disorders. In  
XX situations in which TRIP1 and/or TP2 activity is to be decreased,  
XX such as in cancer cells in which TRIP1 activity and/or TP2 activity  
XX is elevated, TRIP1 and/or TP2 may serve as a target to identify a  
XX molecule which inhibits activity, or which decreases or inhibits  
XX the protein-protein interaction of TRIP1 and TP2, or the binding of  
XX either TRIP1 or TP2 to telomerase RNA. Alternatively, ex vivo or  
XX in vivo gene therapy may be used to administer TRIP1 or TP2  
XX antisense molecules, or DNA constructs may serve to disrupt or  
XX enhance TRIP1 and/or TP2 expression in cells, and to create  
XX dominant negative inhibitors of TRIP1 or TP2.  
XX  
XX Sequence 2627 AA;

Query Match 40.9%; Score 45; DB 19; Length 2627;  
Best Local Similarity 47.1%; Pred. No. 1.5e+02;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 VPSYSFARAHSEVQDL 17  
:||||: :|||  
Db 205 mpsysisgeeevedl 221

RESULT 6  
AAW48586  
ID AAW48586 standard; Protein; 519 AA.  
XX







PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123588.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 04-MAY-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 07-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136382.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145218.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.



PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149802.  
PR 23-AUG-1999; 99US-0149930.  
PR 26-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.

PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 39.1%; Score 43; DB 21; Length 190;  
Best Local Similarity 47.1%; Pred. No. 15;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
  
QY 6 FARHSEVODLINDII 22  
| | : | | | | | | |  
96 fireytskvdldvkdxl 112  
  
RESULT 13  
AAG44996  
ID AAG44996 standard; Protein; 210 AA.  
XX  
AC AAG44996;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Zea mays protein fragment SEQ ID NO: 56434.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence; corn.  
XX  
OS Zea mays subsp. mays.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
  
19-MAY-1999; 99US-0134941.  
20-MAY-1999; 99US-0135124.  
21-MAY-1999; 99US-0135353.  
24-MAY-1999; 99US-0135629.  
25-MAY-1999; 99US-0136021.  
27-MAY-1999; 99US-0136392.  
28-MAY-1999; 99US-0136782.  
01-JUN-1999; 99US-0137222.  
03-JUN-1999; 99US-0137528.  
04-JUN-1999; 99US-0137502.  
07-JUN-1999; 99US-0137724.  
08-JUN-1999; 99US-0138094.  
10-JUN-1999; 99US-0138540.  
10-JUN-1999; 99US-0138847.  
14-JUN-1999; 99US-0139119.  
16-JUN-1999; 99US-0139452.  
16-JUN-1999; 99US-0139453.  
17-JUN-1999; 99US-0139492.  
18-JUN-1999; 99US-0139454.  
18-JUN-1999; 99US-0139455.  
18-JUN-1999; 99US-0139456.  
18-JUN-1999; 99US-0139457.  
18-JUN-1999; 99US-0139458.  
18-JUN-1999; 99US-0139459.  
18-JUN-1999; 99US-0139460.  
18-JUN-1999; 99US-0139461.  
18-JUN-1999; 99US-0139462.  
18-JUN-1999; 99US-0139463.  
18-JUN-1999; 99US-0139750.  
18-JUN-1999; 99US-0139763.  
21-JUN-1999; 99US-0139817.  
22-JUN-1999; 99US-0139899.  
23-JUN-1999; 99US-0140353.  
23-JUN-1999; 99US-0140354.  
28-JUN-1999; 99US-0140823.  
29-JUN-1999; 99US-0140991.  
30-JUN-1999; 99US-0141287.  
01-JUL-1999; 99US-0141842.  
01-JUL-1999; 99US-0142154.  
02-JUL-1999; 99US-0142055.  
06-JUL-1999; 99US-0142390.  
08-JUL-1999; 99US-0142803.  
09-JUL-1999; 99US-0142920.  
12-JUL-1999; 99US-0142977.  
13-JUL-1999; 99US-0143542.  
14-JUL-1999; 99US-0143624.  
15-JUL-1999; 99US-0144005.  
16-JUL-1999; 99US-0144085.  
16-JUL-1999; 99US-0144086.  
19-JUL-1999; 99US-0144325.  
19-JUL-1999; 99US-0144331.  
19-JUL-1999; 99US-0144332.  
19-JUL-1999; 99US-0144333.  
19-JUL-1999; 99US-0144334.  
20-JUL-1999; 99US-0144352.  
20-JUL-1999; 99US-0144632.  
20-JUL-1999; 99US-0144884.  
21-JUL-1999; 99US-0144814.  
21-JUL-1999; 99US-0145086.  
21-JUL-1999; 99US-0145088.  
22-JUL-1999; 99US-0145085.  
22-JUL-1999; 99US-0145087.  
22-JUL-1999; 99US-0145089.  
22-JUL-1999; 99US-0145192.  
23-JUL-1999; 99US-0145145.  
23-JUL-1999; 99US-0145218.  
23-JUL-1999; 99US-0145224.  
26-JUL-1999; 99US-0145276.  
27-JUL-1999; 99US-0145913.  
27-JUL-1999; 99US-0145918.  
27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159233.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 22-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.

PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 26-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 39.1%; Score 43; DB 21; Length 210;  
Best Local Similarity 47.1%; Pred. No. 16;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 6 FARAHSEVQDLIRDI 22  
| | : | : | | : | |  
Db 116 fireytskvddlvkdx 132

RESULT 14  
AAU00297  
ID AAU00297 standard; Protein; 267 AA.  
XX AC AAU00297;  
XX DT 12-SEP-2001 (first entry)  
XX DE Interferon induced polypeptide, IFN5.  
XX KW Interferon induced nucleic acid; autoimmune disease; lupus erythematosus;  
KW immunodeficiency; stroke; acquired immunodeficiency syndrome; AIDS;  
KW graft rejection; viral infection; hepatitis; aplastic anaemia; cancer;  
KW human immunodeficiency virus; HIV; immune-mediated glomerulonephritis;  
KW haematologic disease; chronic neutropenia; myocardial infarction;  
KW neurological disease; Alzheimer's disease; Parkinson's disease; tumour;  
KW amyotrophic lateral sclerosis; spinal muscular atrophy; human; IFN5.  
XX OS Homo sapiens.  
XX WO200118208-A2.  
XX PD 15-MAR-2001.  
XX PF 08-SEP-2000; 2000WO-US24704.  
XX PR 08-SEP-1999; 99US-0152921.  
PR 20-OCT-1999; 99US-0160575.  
PR 20-JAN-2000; 2000US-0177104.  
PR 07-SEP-2000; 2000US-0656633.  
XX (CURA-) CURAGEN CORP.  
PA (BIOJ ) BIOGEN INC.  
XX Payman JA, Da Silva A, Hochman P, Hsu A;  
PA WPI: 2001-235201/24.  
DR N-PSDB; AAS01150.  
XX New interferon induced polypeptides and polynucleotides, useful for the  
PT diagnosis, prevention and treatment of immunological, cell  
PT proliferative disorders, such as lupus erythematosus, cancer, stroke  
PT and Alzheimer's disease  
XX Claim 1; Page 33-35; 134pp; English.  
XX The sequence represents interferon induced polypeptide, IFN5. IFN  
CC nucleic acids and polypeptides are useful for treating or preventing a  
CC pathology associated with IFN polypeptide in a human. they are useful for  
CC determining the presence of or predisposition to a disease associated  
CC with altered levels of IFN polypeptide or polynucleotide. IFN nucleic  
CC acids, polypeptides and antibodies are useful for diagnosis, prevention  
CC or treatment of variety of immunological and cell proliferative  
CC disorders, such as autoimmune diseases e.g. lupus erythematosus,  
CC immunodeficiency diseases such as acquired immunodeficiency syndrome  
CC (AIDS), graft rejection, viral infections including hepatitis and human  
CC immunodeficiency virus (HIV), immune-mediated glomerulonephritis,







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:59:33 ; Search time 87.3 Seconds  
(without alignments)  
5.671 Million cell updates/sec

Title: US-09-290-049a-17

Perfect score: 110  
Sequence: 1 VPSYSFARAHDSVDLIRDI 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

1 number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued\_Patents\_AA:\*
- 1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep.\*
  - 2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep.\*
  - 3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pep.\*
  - 4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pep.\*
  - 5: /cgn2\_6/ptodata/2/iaa/PCRUS.COMB.pep.\*
  - 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100	90.9	1375	4	US-09-210-361-4
2	99	90.0	1475	3	US-09-007-999-2
3	99	90.0	1475	4	US-09-210-361-2
4	62	56.4	1430	3	US-09-008-172-2
5	62	56.4	1430	4	US-09-210-361-6
6	62	56.4	1577	2	US-08-793-824-2
7	45	40.9	2627	2	US-08-751-189-3
8	45	40.9	2627	2	US-09-060-836-3
9	45	40.9	2627	4	US-09-184-445-3
10	43.5	39.5	484	4	US-08-913-578-2
11	43.5	39.5	484	4	US-08-785-427-2
12	43	39.1	639	2	US-08-557-309B-37
13	43	39.1	639	3	US-08-834-306-37
14	43	39.1	639	4	US-08-993-674A-37
15	42	38.2	196	2	US-08-684-024-2
16	42	38.2	196	3	US-08-684-024-9
17	42	38.2	196	3	US-09-145-868-2
18	42	38.2	196	3	US-09-145-868-9
19	41	37.3	933	1	US-08-370-193A-8
20	40	36.4	43	4	US-08-679-006-2
21	40	36.4	1068	3	US-08-390-874C-11
22	40	36.4	1068	4	US-09-265-772-11
23	40	36.4	1069	2	US-08-162-081B-37
24	40	36.4	1069	2	US-08-780-872-37
25	40	36.4	1069	4	US-09-085-957-37
26	40	36.4	1080	2	US-08-162-081B-36
27	40	36.4	1080	2	US-08-780-872-36

28	40	36.4	1080	4	US-09-085-957-36	Sequence 36, Appl
29	40	36.4	1454	4	US-08-392-459-22	Sequence 22, Appl
30	40	36.4	1454	4	US-08-392-459-26	Sequence 26, Appl
31	40	36.4	1454	5	PCT-US91-08525-22	Sequence 22, Appl
32	40	36.4	1454	5	PCT-US91-08525-26	Sequence 26, Appl
33	40	36.4	1454	5	PCT-US93-04384-2	Sequence 2, Appl
34	40	36.4	1454	5	PCT-US93-04384-8	Sequence 8, Appl
35	40	36.4	1454	5	PCT-US93-04384-16	Sequence 16, Appl
36	40	36.4	1454	5	PCT-US93-04384-43	Sequence 43, Appl
37	40	36.4	1454	5	PCT-US93-04384-45	Sequence 45, Appl
38	40	36.4	1454	5	PCT-US93-04384-48	Sequence 48, Appl
39	40	36.4	1497	1	US-08-623-679-7	Sequence 7, Appl
40	40	36.4	1497	3	US-08-933-774-7	Sequence 7, Appl
41	40	36.4	1497	4	US-09-181-030-7	Sequence 9, Appl
42	40	36.4	1533	1	US-08-623-679-9	Sequence 9, Appl
43	40	36.4	1533	3	US-08-933-774-9	Sequence 9, Appl
44	40	36.4	1533	4	US-09-181-030-9	Sequence 9, Appl
45	39	35.5	126	3	US-08-331-035A-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1  
US-09-210-361-4  
; Sequence 4, Application US/09210361  
; Patent No. 6284479  
; GENERAL INFORMATION:  
; APPLICANT: Nicholls, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starches and  
; TITLE OF INVENTION: Latexes in Paper Manufacture  
; FILE REFERENCE: 0357CR  
; CURRENT APPLICATION NUMBER: US/09/210,361  
; CURRENT FILING DATE: 1998-12-11  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/009,620  
; EARLIER FILING DATE: 1998-01-20  
; EARLIER APPLICATION NUMBER: 08/485,243  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/008,172  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSPQ for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 1375  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-210-361-4

Query Match 90.9% Score 100; DB 4; Length 1375;  
Best Local Similarity 90.9%; Pred.No. 1.7e-08;  
Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHDSVDLIRDI 22  
||||| |||||||||  
Db 578 VPSYSFARAHDSVDLIRNI 599

RESULT 2  
US-09-007-999-2  
; Sequence 2, Application US/09007999  
; Patent No. 6087559  
; GENERAL INFORMATION:  
; APPLICANT: Nicholls, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; TITLE OF INVENTION: Latexes in Paper Manufacture  
; FILE REFERENCE: 0356D

; CURRENT APPLICATION NUMBER: US/09/007,999  
; CURRENT FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-007-999-2

Query Match 90.0%; Score 99; DB 3; Length 1475;  
Best Local Similarity 90.9%; Pred. No. 2.8e-08;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 VPSYSFARHDSVDLIRDII 22  
||||| ||||||| |||||  
552 VPSYSFARHDSVDLIRDII 573

RESULT 3  
US-09-210-361-2  
; Sequence 2, Application US/09210361  
; Patent No. 6284479  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starches and  
; FILE REFERENCE: 0357CR  
; CURRENT APPLICATION NUMBER: US/09/210,361  
; CURRENT FILING DATE: 1998-12-11  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/009,620  
; EARLIER FILING DATE: 1998-01-20  
; EARLIER APPLICATION NUMBER: 08/485,243  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/008,172  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-210-361-2

Query Match 90.0%; Score 99; DB 4; Length 1475;  
Best Local Similarity 90.9%; Pred. No. 2.8e-08;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 VPSYSFARHDSVDLIRDII 22  
||||| ||||||| |||||  
552 VPSYSFARHDSVDLIRDII 573

RESULT 4  
US-09-008-172-2  
; Sequence 2, Application US/09008172  
; Patent No. 6127602  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; FILE REFERENCE: 0358D  
; CURRENT APPLICATION NUMBER: US/09/008,172

; CURRENT FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-008-172-2

Query Match 56.4%; Score 62; DB 3; Length 1430;  
Best Local Similarity 65.0%; Pred. No. 0.046;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 SYSFARHDSVDLIRDII 22  
:| | ||||| :| |  
Db 576 NYIFARHDSVQTVIAKII 595

RESULT 5  
US-09-210-361-6  
; Sequence 6, Application US/09210361  
; Patent No. 6284479  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starches and  
; FILE REFERENCE: 0357CR  
; CURRENT APPLICATION NUMBER: US/09/210,361  
; CURRENT FILING DATE: 1998-12-11  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/009,620  
; EARLIER FILING DATE: 1998-01-20  
; EARLIER APPLICATION NUMBER: 08/485,243  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/008,172  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 6  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-210-361-6

Query Match 56.4%; Score 62; DB 4; Length 1430;  
Best Local Similarity 65.0%; Pred. No. 0.046;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 SYSFARHDSVDLIRDII 22  
:| | ||||| :| |  
Db 576 NYIFARHDSVQTVIAKII 595

RESULT 6  
US-08-793-824-2  
; Sequence 2, Application US/08793824  
; Patent No. 5981838  
; GENERAL INFORMATION:  
; APPLICANT: Simeson, Christine Lynn  
; APPLICANT: Giffard, Philip Morrison  
; APPLICANT: Jacques, Nicholas Anthony  
; TITLE OF INVENTION: Genetic Manipulation of Plants to  
; FILE REFERENCE: 0358D  
; CURRENT APPLICATION NUMBER: US/08/008,172

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Griffith Hack & Co  
; STREET: Level 8, 168 Walker Street  
; CITY: No. 5981838th Sydney  
; STATE: New South Wales  
; COUNTRY: Australia  
; ZIP: 2060  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/793,824  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: AU PM7643  
; FILING DATE: 24-AUG-1994  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 61 2 9957 5944  
; TELEFAX: 61 2 957 6288  
; TELEX: 26547  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1577 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Streptococcus salivarius  
US-08-793-824-2

Query Match 56.4%; Score 62; DB 2; Length 1577;  
Best Local Similarity 60.0%; Pred. No. 0.052;  
Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 3 SYSFARHDSVQDLIRDI 22  
Db 661 NYIFVRHDSVQAVLANII 680

RESULT 7  
US-08-751-189-3  
; Sequence 3, Application US/08751189  
; Patent No. 5919656  
; GENERAL INFORMATION:  
; APPLICANT: Harrington, Lea A.  
; TITLE OF INVENTION: No. 5919656el Genes Encoding Telomerase Protein  
; TITLE OF INVENTION: 1  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen, Inc.  
; STREET: 1840 De Havilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/751,189  
; FILING DATE: 15-NOV-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oleski, Nancy A.  
; REGISTRATION NUMBER: 34,688

; REFERENCE/DOCKET NUMBER: A-433  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2627 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-751-189-3

Query Match 40.9%; Score 45; DB 2; Length 2627;  
Best Local Similarity 47.1%; Pred. No. 71;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 1 VPSYSFARHDSVQDL 17  
Db 205 MPYSLSLGEVEEDL 221

RESULT 8  
US-09-060-836-3  
; Sequence 3, Application US/09060836  
; Patent No. 5981707  
; GENERAL INFORMATION:  
; APPLICANT: Harrington, Lea A.  
; TITLE OF INVENTION: No. 5981707el Genes Encoding Telomerase Protein  
; TITLE OF INVENTION: 1  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen, Inc.  
; STREET: 1840 De Havilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/060,836  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/751,189  
; FILING DATE: 15-NOV-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oleski, Nancy A.  
; REGISTRATION NUMBER: 34,688  
; REFERENCE/DOCKET NUMBER: A-433  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2627 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-09-060-836-3

Query Match 40.9%; Score 45; DB 2; Length 2627;  
Best Local Similarity 47.1%; Pred. No. 71;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 1 VPSYSFARHDSVQDL 17  
Db 205 MPYSLSLGEVEEDL 221

RESULT 9

US-09-184-445-3  
; Sequence 3, Application US/09184445  
; Patent No. 6174703  
; GENERAL INFORMATION:  
; APPLICANT: Harrington, Lea A.  
; APPLICANT: Robinson, Murray O.  
; TITLE OF INVENTION: No. 6174703el Genes Encoding Telomerase Protein  
; TITLE OF INVENTION: 1  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen, Inc.  
; STREET: 1840 De Havilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/09/184,445  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/751,189  
; FILING DATE: 15-NOV-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oleski, Nancy A.  
; REGISTRATION NUMBER: 34,688  
; REFERENCE/DOCKET NUMBER: A-433  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2627 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-09-184-445-3

Query Match 40.9%; Score 45; DB 4; Length 2627;  
Best Local Similarity 47.1%; Pred. No. 71;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

1 VPSYSFARHDSVDL 17  
:||||: :|||:  
205 MPSYSLSGEEVEEDL 221

RESULT 10  
US-08-913-578-2  
; Sequence 2, Application US/08913578  
; Patent No. 6218159  
; GENERAL INFORMATION:  
; APPLICANT: Hodgson, John  
; APPLICANT: Lawlor, Elizabeth  
; TITLE OF INVENTION: No. 6218159el tRNA synthetase  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406-0939  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/913,578  
; FILING DATE: 17-JAN-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 9601069.9  
; FILING DATE: 19-JAN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gimmi, Edward R.  
; REGISTRATION NUMBER: 38,891  
; REFERENCE/DOCKET NUMBER: P31352  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-270-4478  
; TELEFAX: 610-270-5090  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 484 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-913-578-2

Query Match 39.5%; Score 43.5; DB 4; Length 484;  
Best Local Similarity 40.9%; Pred. No. 17;  
Matches 9; Conservative 6; Mismatches 4; Indels 3; Gaps 1;

Oy 1 VPSYSFARA---HDSVDLIR 19  
:|:|:| | | :|:|:  
Db 193 IPTYNFAVAIDHMYQISDVIR 214

RESULT 11  
US-08-785-427-2  
; Sequence 2, Application US/08785427  
; Patent No. 6238900  
; GENERAL INFORMATION:  
; APPLICANT: Hodgson, John  
; APPLICANT: Lawlor, Elizabeth  
; TITLE OF INVENTION: No. 6238900el tRNA synthetase  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406-0939  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/785,427  
; FILING DATE: 17-JAN-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 9601069.9  
; FILING DATE: 19-JAN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gimmi, Edward R.  
; REGISTRATION NUMBER: 38,891  
; REFERENCE/DOCKET NUMBER: P31352  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-270-4478  
; TELEFAX: 610-270-5090  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 484 amino acids  
; TYPE: amino acid



FILING DATE: 18-DEC-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.422C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 639 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-993-674A-37

Query Match 39.1%; Score 43; DB 4; Length 639;  
Best Local Similarity 52.9%; Pred. No. 28;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHDSVQDL 17  
Db 479 VPGWSEALLHDAEFQQL 495

RESULT 15  
US-08-684-024-2  
Sequence 2, Application US/08684024  
Patent No. 5834298  
GENERAL INFORMATION:  
APPLICANT: Benezra, Robert  
TITLE OF INVENTION: GENE ENCODING THE HUMAN HOMOLOG OF MAD2  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/684,024  
FILING DATE: 19-JUL-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 1747/46621-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0526  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 196 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-684-024-2

Query Match 38.2%; Score 42; DB 2; Length 196;  
Best Local Similarity 46.7%; Pred. No. 10;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 8 RAHDSVQDLIRDI 22

Db 49 KTHDELKDYIRKIL 63  
Search completed: March 27, 2002, 13:59:33  
Job time: 586 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:20:29 ; Search time 1139.61 Seconds  
(without alignments)  
5.360 Million cell updates/sec

Title: US-09-290-049A-17

Perfect score: 110

Sequence: 1 VPSYSFARAHSEVQDLIRDI 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 3148936 seqs, 277657034 residues

Total number of hits satisfying chosen parameters: 3148936

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_Main:\*

- 1: /cgn2\_6/ptodata/2/paa/PCUS\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/paa/US08\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pep.\*
- 7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pep.\*
- 8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep.\*
- 9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pep.\*
- 10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep.\*
- 11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pep.\*
- 12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pep.\*
- 13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep.\*
- 14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep.\*
- 15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pep.\*
- 16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pep.\*
- 17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pep.\*
- 18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pep.\*
- 19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pep.\*
- 20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pep.\*
- 21: /cgn2\_6/ptodata/2/paa/US097\_COMB.pep.\*
- 22: /cgn2\_6/ptodata/2/paa/US098\_COMB.pep.\*
- 23: /cgn2\_6/ptodata/2/paa/US099\_COMB.pep.\*
- 24: /cgn2\_6/ptodata/2/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	100.0	22	16	US-09-290-049-17
2	108	98.2	22	16	US-09-290-049-19
3	100	90.9	22	16	US-09-290-049-15
4	100	90.9	1375	21	US-09-740-274-4
5	99	90.0	1475	19	US-09-557-848-2
6	99	90.0	1475	21	US-09-740-274-2
7	75.3	68.6	20	16	US-09-290-049-2
8	70	63.6	22	16	US-09-290-049-18
9	68	61.8	2057	18	US-09-499-203-2
					Sequence 17, Appl
					Sequence 19, Appl
					Sequence 15, Appl
					Sequence 4, Appl
					Sequence 2, Appl
					Sequence 2, Appl
					Sequence 2, Appl
					Sequence 18, Appl
					Sequence 2, Appl

10	62	56.4	22	16	US-09-290-049-16	Sequence 16, Appl
11	62	56.4	1430	20	US-09-649-885-2	Sequence 2, Appl
12	62	56.4	1430	21	US-09-740-274-6	Sequence 6, Appl
13	47	42.7	194	18	US-09-417-507-43716	Sequence 43716, A
14	46	41.8	325	24	US-60-324-109-23967	Sequence 23967, A
15	46	41.8	722	1	PCT-US01-08631-58712	Sequence 58712, A
16	45.5	41.4	481	24	US-60-242-578-904	Sequence 904, App
17	45.5	41.4	481	24	US-60-253-625-2248	Sequence 2248, Ap
18	45.5	41.4	481	24	US-60-257-931-3142	Sequence 3142, Ap
19	45.5	41.4	481	24	US-60-269-308-4164	Sequence 4164, Ap
20	45	40.9	214	1	PCT-US01-08631-31600	Sequence 31600, A
21	45	40.9	377	24	US-60-324-109-21938	Sequence 21938, A
22	45	40.9	414	24	US-60-324-109-19998	Sequence 19998, A
23	45	40.9	541	18	US-09-417-507-26458	Sequence 26458, A
24	45	40.9	2627	13	US-08-951-733-3	Sequence 3, Appl
25	44.5	40.5	263	21	US-09-733-089-19824	Sequence 19824, A
26	44.5	40.5	263	22	US-09-816-660-19824	Sequence 19824, A
27	44.5	40.5	501	15	US-09-134-001C-4115	Sequence 4115, Ap
28	44.5	40.5	501	18	US-09-450-969-5378	Sequence 5378, Ap
29	44	40.0	83	24	US-60-207-215-374	Sequence 374, App
30	44	40.0	94	24	US-60-248-798-269	Sequence 269, App
31	44	40.0	129	16	US-09-270-767-32878	Sequence 32878, A
32	44	40.0	129	16	US-09-270-767-48095	Sequence 48095, A
33	44	40.0	129	16	US-09-270-767-188016	Sequence 188016, A
34	44	40.0	129	24	US-60-196-710-6423	Sequence 6423, Ap
35	44	40.0	164	22	US-09-834-366-17052	Sequence 17052, A
36	44	40.0	164	24	US-60-197-873-17052	Sequence 17052, A
37	44	40.0	188	21	US-09-758-460-507	Sequence 507, App
38	44	40.0	208	24	US-60-196-713-3972	Sequence 3972, Ap
39	44	40.0	309	17	US-09-345-473B-37	Sequence 37, Appl
40	44	40.0	309	17	US-09-345-473C-37	Sequence 37, Appl
41	44	40.0	309	20	US-09-609-360B-37	Sequence 37, Appl
42	44	40.0	309	22	US-09-862-027-37	Sequence 37, Appl
43	44	40.0	479	1	PCT-US01-08631-51319	Sequence 51319, A
44	44	40.0	480	7	US-08-350-584-58	Sequence 58, Appl
45	44	40.0	490	1	PCT-US01-03800A-2103	Sequence 2103, Ap

ALIGNMENTS

RESULT 1  
US-09-290-049-17  
; Sequence 17, Application US/09290049  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: FDC98-01p2A  
; CURRENT APPLICATION NUMBER: US/09/290,049  
; CURRENT FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081,550  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/115,142  
; EARLIER FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 17  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. downei  
US-09-290-049-17

Query Match 100.0%; Score 110; DB 16; Length 22;  
Best Local Similarity 100.0%; Pred. No. 4.4e-11;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHSEVQDLIRDI 22  
Db 1 VPSYSFARAHSEVQDLIRDI 22

```

RESULT 2
US-09-290-049-19
; Sequence 19, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CARIES
; FILE REFERENCE: FDC98-012A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 3.0
EQ ID NO 19
LENGTH: 22
; TYPE: PRT
; ORGANISM: S. sobrinus
US-09-290-049-19

Query Match      98.2%; Score 108; DB 16; Length 22;
Best Local Similarity 95.5%; Pred. No. 9.5e-11;
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHDSVQDLIRDI 22
Db 1 VPSYSFARAHDSVQDLIRDI 22

RESULT 3
US-09-290-049-15
; Sequence 15, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CARIES
; FILE REFERENCE: FDC98-012A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 3.0
EQ ID NO 15
LENGTH: 22
; TYPE: PRT
; ORGANISM: S. mutans
US-09-290-049-15

Query Match      90.9%; Score 100; DB 19; Length 1375;
Best Local Similarity 90.9%; Pred. No. 4.5e-07;
Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHDSVQDLIRDI 22
Db 578 VPSYSFARAHDSVQDLIRNI 599

RESULT 5
US-09-557-848-2
; Sequence 2, Application US/09557848
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0356D2
; CURRENT APPLICATION NUMBER: US/09/557,848
; CURRENT FILING DATE: 2000-04-26
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 3.0
EQ ID NO 2
LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-557-848-2

Query Match      90.0%; Score 99; DB 19; Length 1475;
Best Local Similarity 90.9%; Pred. No. 7.3e-07;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHDSVQDLIRDI 22
Db 552 VPSYSFARAHDSVQDLIRDI 573

RESULT 6
US-09-740-274-2
; Sequence 2, Application US/09740274
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper

```

```

US-09-290-049-19
; Sequence 19, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CARIES
; FILE REFERENCE: FDC98-012A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 3.0
EQ ID NO 19
LENGTH: 22
; TYPE: PRT
; ORGANISM: S. sobrinus
US-09-290-049-19

Query Match      98.2%; Score 108; DB 16; Length 22;
Best Local Similarity 95.5%; Pred. No. 9.5e-11;
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHDSVQDLIRDI 22
Db 1 VPSYSFARAHDSVQDLIRDI 22

RESULT 3
US-09-290-049-15
; Sequence 15, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CARIES
; FILE REFERENCE: FDC98-012A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 3.0
EQ ID NO 15
LENGTH: 22
; TYPE: PRT
; ORGANISM: S. mutans
US-09-290-049-15

Query Match      90.9%; Score 100; DB 16; Length 22;
Best Local Similarity 90.9%; Pred. No. 2.1e-09;
Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHDSVQDLIRDI 22
Db 1 VPSYSFARAHDSVQDLIRDI 22

RESULT 4
US-09-740-274-4
; Sequence 4, Application US/09740274
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper

```

```

US-09-290-049-19
; Sequence 19, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CARIES
; FILE REFERENCE: FDC98-012A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 3.0
EQ ID NO 19
LENGTH: 22
; TYPE: PRT
; ORGANISM: S. sobrinus
US-09-290-049-19

Query Match      98.2%; Score 108; DB 16; Length 22;
Best Local Similarity 95.5%; Pred. No. 9.5e-11;
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHDSVQDLIRDI 22
Db 1 VPSYSFARAHDSVQDLIRDI 22

RESULT 3
US-09-290-049-15
; Sequence 15, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CARIES
; FILE REFERENCE: FDC98-012A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 3.0
EQ ID NO 15
LENGTH: 22
; TYPE: PRT
; ORGANISM: S. mutans
US-09-290-049-15

Query Match      90.9%; Score 100; DB 16; Length 22;
Best Local Similarity 90.9%; Pred. No. 2.1e-09;
Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHDSVQDLIRDI 22
Db 1 VPSYSFARAHDSVQDLIRDI 22

RESULT 4
US-09-740-274-4
; Sequence 4, Application US/09740274
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper

```

```

; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans

```

## US-09-740-274-2

```

Query Match          90.0%; Score 99; DB 21; Length 1475;
Best Local Similarity 90.9%; Pred. No. 7.3e-07;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy 1 VPSYSFARAHDSVQDLIRII 22
    ||||| ||||| ||||| |||||
Db 552 VPSYFIRAHDSVQDLIRII 573

```

## RESULT 7

```

US-09-290-049-2
; Sequence 2, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CARRIES
; FILE REFERENCE: FDC98-01p2A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: HDS peptide
US-09-290-049-2

```

```

Query Match          68.6%; Score 75.5; DB 16; Length 20;
Best Local Similarity 89.5%; Pred. No. 2.6e-05;
Matches 17; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```

```

Qy 1 VPSYSFAR-AHDSVQDLI 18
    ||||| ||||| ||||| |||||
Db 1 VPSYFIRAHDSVQDLI 19

```

## RESULT 8

```

US-09-290-049-18
; Sequence 18, Application US/09290049

```

```

; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CARRIES
; FILE REFERENCE: FDC98-01p2A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 22
; TYPE: PRT
; ORGANISM: S. downei
US-09-290-049-18

```

```

Query Match          63.6%; Score 70; DB 16; Length 22;
Best Local Similarity 68.2%; Pred. No. 0.00025;
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

```

```

Qy 1 VPSYSFARAHDSVQDLIRII 22
    ||||| ||||| ||||| |||||
Db 1 VPNYFIRAHDSVQTRIAKII 22

```

## RESULT 9

```

US-09-499-203-2
; Sequence 2, Application US/09499203
; GENERAL INFORMATION:
; APPLICANT: KOSSMANN, Jens
; APPLICANT: WELSH, Thomas
; APPLICANT: QUANZ, Martin
; APPLICANT: KNUTH, Karola
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase
; FILE REFERENCE: 147-196P
; CURRENT APPLICATION NUMBER: US/09/499,203
; CURRENT FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2057
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-499-203-2

```

```

Query Match          61.8%; Score 68; DB 18; Length 2057;
Best Local Similarity 63.2%; Pred. No. 0.19;
Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy 1 VPSYSFARAHDSVQDLIR 19
    :||| |||| : |||
Db 757 IPNYSFVRHVDYDAQDPIR 775

```

## RESULT 10

```

US-09-290-049-16
; Sequence 16, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CARRIES
; FILE REFERENCE: FDC98-01p2A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13

```

; EARLIER APPLICATION NUMBER: 60/115,142  
; EARLIER FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 16  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. mutans  
US-09-290-049-16

Query Match 56.4%; Score 62; DB 16; Length 22;  
Best Local Similarity 65.0%; Pred. No. 0.0055;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 3 SYSFARAHDSVQDLIRDI 22  
: | | | | | | | | | | : | |  
Db 3 NYIFIRAHDSVQTVIAKII 22

RESULT 11  
US-09-649-885-2  
; Sequence 2, Application US/09649885  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; FILE REFERENCE: 035822  
; CURRENT APPLICATION NUMBER: US/09/649,885  
; CURRENT FILING DATE: 2000-08-28  
; PRIOR APPLICATION NUMBER: US 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-649-885-2

Query Match 56.4%; Score 62; DB 20; Length 1430;  
Best Local Similarity 65.0%; Pred. No. 1.2;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 3 SYSFARAHDSVQDLIRDI 22  
: | | | | | | | | | | : | |  
Db 576 NYIFIRAHDSVQTVIAKII 595

RESULT 12  
US-09-740-274-6  
; Sequence 6, Application US/09740274  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16

; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 6  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-740-274-6

Query Match 56.4%; Score 62; DB 21; Length 1430;  
Best Local Similarity 65.0%; Pred. No. 1.2;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 3 SYSFARAHDSVQDLIRDI 22  
: | | | | | | | | | | : | |  
Db 576 NYIFIRAHDSVQTVIAKII 595

RESULT 13  
US-09-417-507-43716  
; Sequence 43716, Application US/09417507  
; GENERAL INFORMATION:  
; APPLICANT: KEITH G. WEINSTOCK ET AL.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ASPERGILLUS  
; FILE REFERENCE: PATH99-10  
; CURRENT APPLICATION NUMBER: US/09/417,507  
; CURRENT FILING DATE: 1999-10-14  
; NUMBER OF SEQ ID NOS: 44312  
; SEQ ID NO 43716  
; LENGTH: 194  
; TYPE: PRT  
; ORGANISM: A.fumigatus  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (5),(8),(56)  
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are un  
US-09-417-507-43716

Query Match 42.7%; Score 47; DB 18; Length 194;  
Best Local Similarity 42.1%; Pred. No. 32;  
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 2 PSYSFARAHDSVQDLIRD 20  
: | | | | | : | | : | |  
Db 14 PSFSFRSRSKEVREAMRN 32

RESULT 14  
US-60-324-109-23967  
; Sequence 23967, Application US/60324109  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Edgerton, Michael D  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Stein, Joshua  
; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT  
; FILE REFERENCE: 38-10(52726)B  
; CURRENT APPLICATION NUMBER: US/60/324,109  
; CURRENT FILING DATE: 2001-09-21  
; NUMBER OF SEQ ID NOS: 33196  
; SEQ ID NO 23967  
; LENGTH: 325  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
US-60-324-109-23967

Query Match 41.8%; Score 46; DB 24; Length 325;  
Best Local Similarity 50.0%; Pred. No. 91;  
Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 1 VPSYFARAHDSVQDLIRD 20  
||| ||| : ||| : |  
Db 50 VPLYITAREHNGHVQLLVAD 69

## RESULT 15

PCT-US01-08631-58712  
; Sequence 58712, Application PC/TUS0108631  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 21272-049  
; CURRENT APPLICATION NUMBER: PCT/US01/08631  
; CURRENT FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 09/540,217  
; PRIOR FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 09/649,167  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 60736  
; SOFTWARE: Custom  
; SEQ ID NO 58712  
; LENGTH: 722  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (67)..(97)  
; OTHER INFORMATION: kw TRANSCRIPTASE REVERSE II ORF2 domain identified by  
; OTHER INFORMATION: eMATRIX, accession number DM01354R, p-value=4.263e-16, raw score  
; OTHER INFORMATION: 8.50  
PCT-US01-08631-58712

Query Match 41.8%; Score 46; DB 1; Length 722;  
Best Local Similarity 47.4%; Pred. No. 2.6e+02;  
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 3 SYSPARAHDSVQDLIRDI 21  
:|:| | ||: |||  
Db 7 TYTFRLHQEEVESLNRP I 25

Search completed: March 27, 2002, 14:20:30  
time: 1577 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:22:49 ; Search time 137.48 seconds  
(without alignments)  
11.042 Million cell updates/sec

Title: US-09-290-049A-17  
Perfect score: 110  
Sequence: 1 VPSYFARAHDSVDLIRDI 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 332938 seqs, 6899538 residues

Total number of hits satisfying chosen parameters: 332938

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep1.\*  
7: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep.\*  
8: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	110	100.0	22	6	US-09-562-328-30
2	110	100.0	22	6	US-09-290-049A-17
3	108	98.2	22	6	US-09-562-328-32
4	108	98.2	22	6	US-09-290-049A-19
5	100	90.9	22	6	US-09-562-328-28
6	100	90.9	22	6	US-09-290-049A-15
7	86	78.2	19	6	US-09-562-328-27
8	86	78.2	19	6	US-09-290-049A-2
9	86	78.2	23	6	US-09-562-328-44
10	79	71.8	523	6	US-09-604-957-5
11	75.5	68.6	20	6	US-09-562-328-26
12	70	63.6	22	6	US-09-562-328-31
13	70	63.6	22	6	US-09-290-049A-18
14	68	61.8	535	6	US-09-604-957-7
15	68	61.8	1278	6	US-09-604-957-3
16	63	57.3	584	6	US-09-604-957-6
17	62	56.4	22	6	US-09-562-328-29
18	62	56.4	22	6	US-09-290-049A-16
19	62	56.4	545	6	US-09-604-957-4
20	53	48.2	215	6	US-09-675-784A-8997
21	45.5	41.4	481	1	PCT-US02-03987-5584
22	45.5	41.4	481	6	US-09-815-242-5584
23	45.5	41.4	481	7	US-10-072-851-5584
24	45.5	41.4	487	1	PCT-US02-03987-12456
25	45.5	41.4	487	6	US-09-815-242-12456

26	45.5	41.4	487	7	US-10-072-851-12456	Sequence 12456, A
27	45	40.9	274	6	US-09-675-784A-9123	Sequence 9123, Ap
28	44	40.0	309	6	US-09-609-360C-37	Sequence 37, Appl
29	44	40.0	309	6	US-09-345-473E-37	Sequence 37, Appl
30	44	40.0	521	6	US-09-646-673A-117	Sequence 117, App
31	44	40.0	564	6	US-09-646-673A-180	Sequence 180, App
32	44	40.0	2193	8	US-60-338-690-5	Sequence 5, Appl
33	44	39.1	188	6	US-09-675-784A-10453	Sequence 10453, A
34	43	39.1	267	6	US-09-656-633-10	Sequence 10, Appl
35	43	39.1	271	6	US-09-656-633-14	Sequence 14, Appl
36	43	39.1	304	6	US-09-656-633-8	Sequence 8, Appl
37	43	39.1	308	6	US-09-656-633-12	Sequence 12, Appl
38	42.5	38.6	486	1	PCT-US02-03987-13455	Sequence 13455, A
39	42.5	38.6	486	6	US-09-815-242-13455	Sequence 13455, A
40	42.5	38.6	486	7	US-10-072-851-13455	Sequence 13455, A
41	42	38.2	219	7	US-10-015-127-10718	Sequence 10718, A
42	42	38.2	227	5	US-09-993-308-6	Sequence 6, Appl
43	42	38.2	304	6	US-09-708-427-7240	Sequence 7240, Ap
44	42	38.2	323	6	US-09-708-427-7239	Sequence 7239, Ap
45	42	38.2	353	6	US-09-614-150-4167	Sequence 4167, Ap

## ALIGNMENTS

RESULT 1

US-09-562-328-30  
; Sequence 30, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARRIES  
; FILE REFERENCE: 04995-0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 30  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-30

Query Match 100.0%; Score 110; DB 6; Length 22;  
Best Local Similarity 100.0%; Pred. No. 4e-11;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 VPSYFARAHDSVDLIRDI 22  
Db 1 VPSYFARAHDSVDLIRDI 22  
|||||

RESULT 2

US-09-290-049A-17  
; Sequence 17, Application US/09290049A  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL CARRIES  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 17  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. downei  
US-09-290-049A-17

Query Match 100.0%; Score 110; DB 6; Length 22;  
Best Local Similarity 100.0%; Pred. No. 4e-11;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPSYSFARAHDSVQDLIRDI 22  
| | | | | | | | | | | | | | | | | | | | | |  
DB 1 VPSYSFARAHDSVQDLIRDI 22

RESULT 3  
US-09-562-328-32  
; Sequence 32, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 32  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-32

Query Match 98.2%; Score 108; DB 6; Length 22;  
Best Local Similarity 95.5%; Pred. No. 8.4e-11;  
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPSYSFARAHDSVQDLIRDI 22  
| | | | | | | | | | | | | | | | | | | | | |  
DB 1 VPSYSFARAHDSVQDLIRDI 22

RESULT 4  
US-09-290-049A-19  
; Sequence 19, Application US/09290049A  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 19  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. sobrinus  
US-09-290-049A-19

Query Match 98.2%; Score 108; DB 6; Length 22;  
Best Local Similarity 95.5%; Pred. No. 8.4e-11;

Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPSYSFARAHDSVQDLIRDI 22  
| | | | | | | | | | | | | | | | | | | | | |  
DB 1 VPSYSFARAHDSVQDLIRDI 22

RESULT 5  
US-09-562-328-28  
; Sequence 28, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 28  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-28

Query Match 90.9%; Score 100; DB 6; Length 22;  
Best Local Similarity 90.9%; Pred. No. 1.6e-09;  
Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSFARAHDSVQDLIRDI 22  
| | | | | | | | | | | | | | | | | | | | | |  
DB 1 VPSYSFARAHDSVQDLIRDI 22

RESULT 6  
US-09-290-049A-15  
; Sequence 15, Application US/09290049A  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. mutans  
US-09-290-049A-15

Query Match 90.9%; Score 100; DB 6; Length 22;  
Best Local Similarity 90.9%; Pred. No. 1.6e-09;  
Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSFARAHDSVQDLIRDI 22  
| | | | | | | | | | | | | | | | | | | | | |  
DB 1 VPSYSFARAHDSVQDLIRDI 22

RESULT 7  
US-09-562-328-27



```
; Sequence 27, Application US/09562328
; GENERAL INFORMATION:
; APPLICANT: LEES, ANDREW
; APPLICANT: TAUBMAN, MARTIN A.
; APPLICANT: SMITH, DANIEL J.
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES
; FILE REFERENCE: 04995.0046-01
; CURRENT APPLICATION NUMBER: US/09/562,328
; CURRENT FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 09/288,965
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Streptococcus sp.
US-09-562-328-27

Query Match      78.2%; Score 86; DB 6; Length 19;
Best Local Similarity 94.4%; Pred. No. 2.2e-07;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSFARAHSEVQDLI 18
   ||||| ||||| |||||
Db 1 VPSYSFIRAHSEVQDLI 18

RESULT 8
US-09-290-049a-2
; Sequence 2, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290,049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: HDS peptide
US-09-290-049a-2

Query Match      78.2%; Score 86; DB 6; Length 19;
Best Local Similarity 94.4%; Pred. No. 2.2e-07;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSFARAHSEVQDLI 18
   ||||| ||||| |||||
Db 1 VPSYSFIRAHSEVQDLI 18

RESULT 9
US-09-562-328-44
; Sequence 44, Application US/09562328
; GENERAL INFORMATION:
; APPLICANT: LEES, ANDREW
; APPLICANT: TAUBMAN, MARTIN A.
; APPLICANT: SMITH, DANIEL J.
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES
; FILE REFERENCE: 04995.0046-01
```

```
; CURRENT APPLICATION NUMBER: US/09/562,328
; CURRENT FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 09/288,965
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 44
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Streptococcus sp.
US-09-562-328-44

Query Match      78.2%; Score 86; DB 6; Length 23;
Best Local Similarity 94.4%; Pred. No. 2.8e-07;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSFARAHSEVQDLI 18
   ||||| ||||| |||||
Db 5 VPSYSFIRAHSEVQDLI 22

RESULT 10
US-09-604-957-5
; Sequence 5, Application US/09604957
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-604-957-5

Query Match      71.8%; Score 79; DB 6; Length 523;
Best Local Similarity 63.6%; Pred. No. 0.00016;
Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPSYSFARAHSEVQDLIRII 22
   ||||| ||||| |||||
Db 146 IPNYSFVRANDSEVQTIVIAQIV 167

RESULT 11
US-09-562-328-2
; Sequence 26, Application US/09562328
; GENERAL INFORMATION:
; APPLICANT: LEES, ANDREW
; APPLICANT: TAUBMAN, MARTIN A.
; APPLICANT: SMITH, DANIEL J.
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES
; FILE REFERENCE: 04995.0046-01
; CURRENT APPLICATION NUMBER: US/09/562,328
; CURRENT FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 09/288,965
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Streptococcus sp.
```

US-09-562-328-26

Query Match 68.6%; Score 75.5; DB 6; Length 20;  
Best Local Similarity 89.5%; Pred. No. 1.1e-05;  
Matches 17; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1 VPSYSFAR-AHDSEVDLII 18  
| | | | | | | | | | | | | | | | | |  
Db 1 VPSYSFIRAHDSVDLII 19

RESULT 12  
US-09-562-328-31  
; Sequence 31, Application US/09562328  
; GENERAL INFORMATION:

; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 31  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-31

Query Match 63.6%; Score 70; DB 6; Length 22;  
Best Local Similarity 68.2%; Pred. No. 9.4e-05;  
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHDSVDLIRDII 22  
| | | | | | | | | | | | | | | | | |  
Db 1 VPNYFIRAHDSVQTRIAKII 22

RESULT 13  
US-09-290-049A-18  
; Sequence 18, Application US/09290049A  
; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CARIES  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 18  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. downei  
US-09-290-049A-18

Query Match 63.6%; Score 70; DB 6; Length 22;  
Best Local Similarity 68.2%; Pred. No. 9.4e-05;  
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHDSVDLIRDII 22  
| | | | | | | | | | | | | | | | | |

Db 1 VPNYFIRAHDSVQTRIAKII 22

RESULT 14  
US-09-604-957-7  
; Sequence 7, Application US/09604957  
; GENERAL INFORMATION:

; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
; APPLICANT: DIJKHUIZEN, LUBBERT  
; APPLICANT: RAHAOUI, HAKIM  
; APPLICANT: LEER, ROBERT-JAN  
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
; FILE REFERENCE: BO 43388  
; CURRENT APPLICATION NUMBER: US/09/604,957  
; CURRENT FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 00201871.1  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 535  
; TYPE: PRT  
; ORGANISM: Lactobacillus reuteri  
US-09-604-957-7

Query Match 61.8%; Score 68; DB 6; Length 535;  
Best Local Similarity 54.5%; Pred. No. 0.0095;  
Matches 12; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHDSVDLIRDII 22  
| | | | | | | | | | | | | | | | | |  
Db 144 IPNYSFVRAHDNNSQDQIONAI 165

RESULT 15  
US-09-604-957-3  
; Sequence 3, Application US/09604957  
; GENERAL INFORMATION:

; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
; APPLICANT: DIJKHUIZEN, LUBBERT  
; APPLICANT: RAHAOUI, HAKIM  
; APPLICANT: LEER, ROBERT-JAN  
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
; FILE REFERENCE: BO 43388  
; CURRENT APPLICATION NUMBER: US/09/604,957  
; CURRENT FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 00201871.1  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 1278  
; TYPE: PRT  
; ORGANISM: Lactobacillus reuteri  
US-09-604-957-3

Query Match 61.8%; Score 68; DB 6; Length 1278;  
Best Local Similarity 54.5%; Pred. No. 0.028;  
Matches 12; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 1 VPSYSFARAHDSVDLIRDII 22  
| | | | | | | | | | | | | | | | | |  
Db 620 IPNYSFVRAHDNNSQDQIONAI 641

Search completed: March 27, 2002, 14:22:49  
Job time: 1696 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:59:32 ; Search time 87.3 Seconds  
(without alignments)  
5.671 Million cell updates/sec

Title: US-09-290-049A-16

Perfect score: 107

Sequence: 1 MANYIFIRAHDSVOTVIKII 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued\_Patents\_AA:\*
- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/iaa/PCRTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107	100.0	1430	3	US-09-008-172-2
2	107	100.0	1430	4	US-09-210-361-6
3	89	83.2	1577	2	US-08-793-824-2
4	72	67.3	1475	3	US-09-007-999-2
5	72	67.3	1475	4	US-09-210-361-2
6	68	63.6	1375	4	US-09-210-361-4
7	39	36.4	182	4	US-09-338-907-133
8	39	36.4	185	4	US-09-338-907-136
9	39	36.4	228	4	US-09-338-907-70
10	39	36.4	300	4	US-09-338-907-135
11	39	36.4	315	4	US-09-338-907-134
12	39	36.4	353	2	US-08-996-306-4
13	39	36.4	353	4	US-09-338-907-4
14	39	36.4	364	2	US-08-996-306-5
15	39	36.4	364	4	US-09-338-907-5
16	38	35.5	146	2	US-08-647-960-8
17	38	35.5	257	4	US-09-287-097-2
18	38	35.5	260	2	US-08-685-992-4
19	38	35.5	260	2	US-08-685-992-25
20	38	35.5	260	2	US-09-144-925-4
21	38	35.5	260	2	US-09-144-925-25
22	38	35.5	261	2	US-08-685-992-3
23	38	35.5	261	2	US-09-144-925-3
24	38	35.5	1091	3	US-08-986-485-5
25	38	35.5	2308	1	US-08-015-973-1
26	38	35.5	2308	2	US-08-448-164-1
27	38	35.5	2308	4	US-08-081-929-2

28	37.5	35.0	613	1	US-08-405-615-1	Sequence 1, Appli
29	37.5	35.0	613	2	US-08-461-234-1	Sequence 1, Appli
30	37.5	35.0	613	2	US-08-463-480-1	Sequence 1, Appli
31	37.5	35.0	614	1	US-08-225-224-1	Sequence 1, Appli
32	37.5	35.0	614	3	US-08-722-258-1	Sequence 1, Appli
33	37.5	35.0	614	5	PCT-US95-04468-1	Sequence 1, Appli
34	37.5	35.0	635	4	US-09-046-992-2	Sequence 2, Appli
35	37.5	35.0	638	3	US-09-047-148-2	Sequence 2, Appli
36	37	34.6	84	3	US-09-013-067A-4	Sequence 4, Appli
37	37	34.6	112	4	US-08-899-330-9	Sequence 9, Appli
38	37	34.6	196	2	US-08-684-024-2	Sequence 2, Appli
39	37	34.6	196	2	US-08-684-024-9	Sequence 9, Appli
40	37	34.6	196	3	US-09-145-868-2	Sequence 2, Appli
41	37	34.6	196	3	US-09-145-868-9	Sequence 9, Appli
42	37	34.6	221	3	US-09-013-067A-8	Sequence 8, Appli
43	37	34.6	364	1	US-08-650-275-4	Sequence 4, Appli
44	37	34.6	364	3	US-09-181-318-4	Sequence 4, Appli
45	37	34.6	876	1	US-08-785-429-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-09-008-172-2  
; Sequence 2, Application US/09008172  
; Patent No. 6127602  
; GENERAL INFORMATION:  
; APPLICANT: Nicholls, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; FILE OF INVENTION: Latexes in Paper Manufacture  
; FILE REFERENCE: 0358D  
; CURRENT APPLICATION NUMBER: US/09/008,172  
; CURRENT FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-008-172-2

Query Match 100.0%; Score 107; DB 3; Length 1430;  
Best Local Similarity 100.0%; Pred. No. 8.6e-11;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MANYIFIRAHDSVOTVIKII 22  
Db 574 MANYIFIRAHDSVOTVIKII 595  
RESULT 2  
US-09-210-361-6  
; Sequence 6, Application US/09210361  
; Patent No. 6284479  
; GENERAL INFORMATION:  
; APPLICANT: Nicholls, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starches and  
; FILE OF INVENTION: Latexes in Paper Manufacture  
; FILE REFERENCE: 0357CR  
; CURRENT APPLICATION NUMBER: US/09/210,361  
; CURRENT FILING DATE: 1998-12-11  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/009,620  
; EARLIER FILING DATE: 1998-01-20  
; EARLIER APPLICATION NUMBER: 08/485,243  
; EARLIER FILING DATE: 1995-06-07

EARLIER APPLICATION NUMBER: 09/008,172  
EARLIER FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/482,711  
EARLIER FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSEQ for Windows Version 3.0  
SEQ ID NO 6  
LENGTH: 1430  
TYPE: PRT  
ORGANISM: streptococcus mutans  
US-09-210-361-2

Query Match 100.0%; Score 107; DB 4; Length 1430;  
Best Local Similarity 100.0%; Pred. No. 8.6e-11;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MANYIFIRAHDSVQTVIAKII 22  
|||||  
574 MANYIFIRAHDSVQTVIAKII 595

RESULT 3  
US-08-793-824-2  
Sequence 2, Application US/08793824  
Patent No. 5981838  
GENERAL INFORMATION:  
APPLICANT: Simpson, Christine Lynn  
APPLICANT: Giffard, Philip Morrison  
APPLICANT: Jacques, Nicholas Anthony  
TITLE OF INVENTION: Genetic Manipulation of Plants to  
INCREASE STORED CARBOHYDRATES  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Griffith Hack & Co  
STREET: Level 8, 168 Walker Street  
CITY: No. 5981838th Sydney  
STATE: New South Wales  
COUNTRY: Australia  
ZIP: 2060  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/793,824  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: AU PM7643  
FILING DATE: 24-AUG-1994  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 61 2 957 5944  
TELEFAX: 61 2 957 6288  
TELEX: 26547  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1577 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus salivarius  
US-08-793-824-2

Query Match 83.2%; Score 89; DB 2; Length 1577;  
Best Local Similarity 81.0%; Pred. No. 1.7e-07;  
Matches 17; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ANYIFIRAHDSVQTVIAKII 22  
:| ||||| :|||

Db 660 ANYIFVRAHDSVQAVLANII 680  
||||| :|||

RESULT 4  
US-09-007-999-2  
Sequence 2, Application US/09007999  
Patent No. 6087559  
GENERAL INFORMATION:  
APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Substitutes for Modified Starch and  
LATEXES IN PAPER MANUFACTURE  
FILE REFERENCE: 0356D  
CURRENT APPLICATION NUMBER: US/09/007,999  
CURRENT FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/478,704  
EARLIER FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: FastSEQ for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 1475  
TYPE: PRT  
ORGANISM: Streptococcus mutans  
US-09-007-999-2

Query Match 67.3%; Score 72; DB 3; Length 1475;  
Best Local Similarity 75.0%; Pred. No. 0.00018;  
Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 3 NYIFIRAHDSVQTVIAKII 22  
:| ||||| :|||

Db 554 SYSFIRAHDSVQDLIADII 573

RESULT 5  
US-09-210-361-2  
Sequence 2, Application US/09210361  
Patent No. 6284479  
GENERAL INFORMATION:  
APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Substitutes for Modified Starches and  
LATEXES IN PAPER MANUFACTURE  
FILE REFERENCE: 0357CR  
CURRENT APPLICATION NUMBER: US/09/210,361  
CURRENT FILING DATE: 1998-12-11  
EARLIER APPLICATION NUMBER: 09/007,999  
EARLIER FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/478,704  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 09/009,620  
EARLIER FILING DATE: 1998-01-20  
EARLIER APPLICATION NUMBER: 08/485,243  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 09/008,172  
EARLIER FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/482,711  
EARLIER FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSEQ for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 1475  
TYPE: PRT  
ORGANISM: Streptococcus mutans  
US-09-210-361-2

Query Match 67.3%; Score 72; DB 4; Length 1475;  
Best Local Similarity 75.0%; Pred. No. 0.00018;  
Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 3 NYIFIRAHDSVQTVIAKII 22  
:| ||||| :|||

```

RESULT          9
US-09-338-907-70
; Sequence 70, Application US/09338907
; Patent No. 6265546
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Ilya, Chumakov
; APPLICANT: Bouguerelet, Lydie
; TITLE OF INVENTION: PROSTATE CANCER GENE
; FILE REFERENCE: GENSET.18C1PC
; CURRENT APPLICATION NUMBER: US/09/338,907
; CURRENT FILING DATE: 1999-06-23
; EARLIER APPLICATION NUMBER: 08/996,306
; EARLIER FILING DATE: 1997-12-22
; EARLIER APPLICATION NUMBER: 60/099,658

```

; EARLIER FILING DATE: 1998-09-09  
; EARLIER APPLICATION NUMBER: 09/218,207  
; EARLIER FILING DATE: 1998-12-22  
; NUMBER OF SEQ ID NOS: 578  
; SOFTWARE: Patent.pm  
; SEQ ID NO 70  
; LENGTH: 228  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-338-907-70

Query Match 36.4%; Score 39; DB 4; Length 228;  
Best Local Similarity 35.0%; Pred. No. 18;  
Matches 7; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 3 NYIFRAHDSVQTVIAKII 22  
| | : | | : | | :  
75 NIIYLANHQSTVDWIVADIL 94

RESULT 10  
US-09-338-907-135  
; Sequence 135, Application US/09338907  
; Patent No. 6265546  
; GENERAL INFORMATION:  
; APPLICANT: Cohen, Daniel  
; APPLICANT: Blumenfeld, Marta  
; APPLICANT: Ilyia, Chumakov  
; TITLE OF INVENTION: PROSTATE CANCER GENE  
; FILE REFERENCE: GENSET-18CPICP  
; CURRENT APPLICATION NUMBER: US/09/338,907  
; CURRENT FILING DATE: 1999-06-23  
; EARLIER APPLICATION NUMBER: 08/996,306  
; EARLIER FILING DATE: 1997-12-22  
; EARLIER APPLICATION NUMBER: 60/099,658  
; EARLIER FILING DATE: 1998-09-09  
; EARLIER APPLICATION NUMBER: 09/218,207  
; EARLIER FILING DATE: 1998-12-22  
; NUMBER OF SEQ ID NOS: 578  
; SOFTWARE: Patent.pm  
; SEQ ID NO 135  
; LENGTH: 300  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: 81..83  
; OTHER INFORMATION: Box I  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: 160..165  
; OTHER INFORMATION: Box II  
US-09-338-907-135

Query Match 36.4%; Score 39; DB 4; Length 300;  
Best Local Similarity 35.0%; Pred. No. 24;  
Matches 7; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 3 NYIFRAHDSVQTVIAKII 22  
| | : | | : | | :  
75 NIIYLANHQSTVDWIVADIL 94

RESULT 11  
US-09-338-907-134  
; Sequence 134, Application US/09338907  
; Patent No. 6265546  
; GENERAL INFORMATION:  
; APPLICANT: Cohen, Daniel  
; APPLICANT: Blumenfeld, Marta

; APPLICANT: Ilyia, Chumakov  
; TITLE OF INVENTION: PROSTATE CANCER GENE  
; FILE REFERENCE: GENSET-18CPICP  
; CURRENT APPLICATION NUMBER: US/09/338,907  
; CURRENT FILING DATE: 1999-06-23  
; EARLIER APPLICATION NUMBER: 08/996,306  
; EARLIER FILING DATE: 1997-12-22  
; EARLIER APPLICATION NUMBER: 60/099,658  
; EARLIER FILING DATE: 1998-09-09  
; EARLIER APPLICATION NUMBER: 09/218,207  
; EARLIER FILING DATE: 1998-12-22  
; NUMBER OF SEQ ID NOS: 578  
; SOFTWARE: Patent.pm  
; SEQ ID NO 134  
; LENGTH: 315  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: 81..83  
; OTHER INFORMATION: Box I  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: 160..165  
; OTHER INFORMATION: Box II  
US-09-338-907-134

Query Match 36.4%; Score 39; DB 4; Length 315;  
Best Local Similarity 35.0%; Pred. No. 26;  
Matches 7; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 3 NYIFRAHDSVQTVIAKII 22  
| | : | | : | | :  
75 NIIYLANHQSTVDWIVADIL 94

RESULT 12  
US-08-996-306-4  
; Sequence 4, Application US/08996306  
; Patent No. 5945522  
; GENERAL INFORMATION:  
; APPLICANT: Cohen, Daniel  
; APPLICANT: Chumakov, Ilyia  
; APPLICANT: Blumenfeld, Marta  
; APPLICANT: Bougueleret, Lydie  
; TITLE OF INVENTION: Prostate cancer gene  
; NUMBER OF SEQUENCES: 68  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 501 West Broadway  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92101-3505  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy Disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: Win95  
; SOFTWARE: Word  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/996,306  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Israel, Ned A.  
; REGISTRATION NUMBER: 29,655  
; REFERENCE/DOCKET NUMBER: GENSET.018A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 235-8550  
; TELEFAX: (619) 235-0176  
; INFORMATION FOR SEQ ID NO: 4:





```
; SEQ ID NO 4
; LENGTH: 353
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: HELIX
; LOCATION: 1..33
; OTHER INFORMATION: Rao and Argos identification method, potential helix
; FEATURE:
; NAME/KEY: HELIX
; LOCATION: 4..20
; OTHER INFORMATION: Klein, Kanehisa and Delisi identification method, potential
; FEATURE:
; NAME/KEY: HELIX
; LOCATION: 4..24
; OTHER INFORMATION: Eisenberg, Schwarz, Komarony, Wall identification method, P
; FEATURE:
; NAME/KEY: MYRISTATE
; LOCATION: 12..16
; OTHER INFORMATION: Prosite match
; FEATURE:
; NAME/KEY: HELIX
; LOCATION: 50..70
; OTHER INFORMATION: Eisenberg, Schwarz, Komarony, Wall identification method, P
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 57..59
; OTHER INFORMATION: Prosite match
; FEATURE:
; NAME/KEY: HELIX
; LOCATION: 76..96
; OTHER INFORMATION: Eisenberg, Schwarz, Komarony, Wall identification method, P
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 78
; OTHER INFORMATION: potential Tyrosine kinase site, Prosite match
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 84
; OTHER INFORMATION: potential caseine kinase II site, Prosite match
; FEATURE:
; NAME/KEY: SITE
; LOCATION: 94..115
; OTHER INFORMATION: potential Leucine zipper site, Prosite match
; FEATURE:
; NAME/KEY: MYRISTATE
; LOCATION: 119..123
; OTHER INFORMATION: potential site, Prosite match
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 133
; OTHER INFORMATION: potential protein kinase C, Prosite match
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 147
; OTHER INFORMATION: potential caseine kinase II site, Prosite match
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 194
; OTHER INFORMATION: potential protein kinase C, Prosite match
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 215
; OTHER INFORMATION: potential Tyrosine kinase site, Prosite match
; FEATURE:
; NAME/KEY: SULFATATION
; LOCATION: 221
; OTHER INFORMATION: Prosite match
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 233
; OTHER INFORMATION: potential cAMP and cGMP dependant protein kinase site, P
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 235
; OTHER INFORMATION: potential caseine kinase II site, Prosite match
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 306
; OTHER INFORMATION: potential protein kinase C, Prosite match
; FEATURE:
; NAME/KEY: HELIX
; LOCATION: 310..330
; OTHER INFORMATION: Eisenberg, Schwarz, Komarony, Wall identification method,
; FEATURE:
; NAME/KEY: MYRISTATE
; LOCATION: 319..323
; OTHER INFORMATION: Prosite match
; FEATURE:
; NAME/KEY: MYRISTATE
; LOCATION: 323..327
; OTHER INFORMATION: Prosite match
; FEATURE:
; NAME/KEY: AMIDATION
; LOCATION: 329
; OTHER INFORMATION: Prosite match
; NAME/KEY: HELIX
; LOCATION: 333..353
; OTHER INFORMATION: Eisenberg, Schwarz, Komarony, Wall identification method,
; FEATURE:
; NAME/KEY: MYRISTATE
; LOCATION: 341..345
; OTHER INFORMATION: Prosite match
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 350
; OTHER INFORMATION: potential protein kinase C, Prosite match
; US-09-338-907-4

Query Match          36.4%; Score 39; DB 4; Length 353;
Best Local Similarity 35.0%; Pred. No. 30;
Matches 7; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Oy 3 NVIFIRAHDSVQTVIAKII 22
   | | | | | | | | | |
Db 75 NIIYLANHQSTVDWIVADIL 94

RESULT 14
US-08-996-306-5
; Sequence 5, Application US/08996306
; Patent No. 5945522
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Chumakov, Ilya
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Prostate cancer gene
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/996,306
; FILING DATE:
```

Search completed: March 27, 2002, 13:59:33  
Job time: 586 sec



Result No.	Score	Query Match	Length	DB	ID	Description
1	107	100.0	22	16	US-09-290-049-16	Sequence 16, Appl
2	107	100.0	1430	20	US-09-649-885-2	Sequence 2, Appli
3	107	100.0	1430	21	US-09-740-274-6	Sequence 6, Appli
4	90	84.1	22	16	US-09-290-049-18	Sequence 18, Appl
5	72	67.3	1475	19	US-09-557-848-2	Sequence 2, Appli
6	72	67.3	1475	21	US-09-740-274-2	Sequence 2, Appli
7	68	63.6	22	16	US-09-290-049-15	Sequence 15, Appl
8	68	63.6	1375	21	US-09-740-274-4	Sequence 4, Appli
9	64	59.8	22	16	US-09-290-049-19	Sequence 19, Appl

RESULT 2  
US-09-649-885-2  
; Sequence 2, Application US/09649885  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; FILE REFERENCE: 035802  
; CURRENT APPLICATION NUMBER: US/09/649,885  
; CURRENT FILING DATE: 2000-08-28  
; PRIOR APPLICATION NUMBER: US 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 1430  
TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-649-885-2

Query Match 100.0%; Score 107; DB 20; Length 1430;  
Best Local Similarity 100.0%; Pred. No. 8.3e-09;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MANYIFIRAHDSVQTVIAKII 22  
|||||  
DB 574 MANYIFIRAHDSVQTVIAKII 595

RESULT 3  
US-09-740-274-6  
; Sequence 6, Application US/09740274  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CHD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 6  
LENGTH: 1430  
TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-740-274-6

Query Match 100.0%; Score 107; DB 21; Length 1430;  
Best Local Similarity 100.0%; Pred. No. 8.3e-09;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MANYIFIRAHDSVQTVIAKII 22  
|||||  
DB 574 MANYIFIRAHDSVQTVIAKII 595

RESULT 4  
US-09-290-049-18  
; Sequence 18, Application US/09290049  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: FDC98-01P2A  
; CURRENT APPLICATION NUMBER: US/09/290,049  
; CURRENT FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081,550  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/115,142  
; EARLIER FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 18  
LENGTH: 22  
TYPE: PRT  
; ORGANISM: S. downei  
US-09-290-049-18

Query Match 84.1%; Score 90; DB 16; Length 22;  
Best Local Similarity 90.0%; Pred. No. 5.1e-08;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NYFIRAHDSVQTVIAKII 22  
|||||  
DB 3 NYFIRAHDSVQTVIAKII 22

RESULT 5  
US-09-557-848-2  
; Sequence 2, Application US/09557848  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; FILE REFERENCE: 0356D2  
; CURRENT APPLICATION NUMBER: US/09/557,848  
; CURRENT FILING DATE: 2000-04-26  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 1475  
TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-557-848-2

Query Match 67.3%; Score 72; DB 19; Length 1475;  
Best Local Similarity 75.0%; Pred. No. 0.0096;  
Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 NYFIRAHDSVQTVIAKII 22  
:|||||  
DB 554 SYSFIRAHDSVQDLIADII 573

RESULT 6  
US-09-740-274-2  
; Sequence 2, Application US/09740274  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD

```
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2
```

```
Query Match 67.3%; Score 72; DB 21; Length 1475;
Best Local Similarity 75.0%; Pred. No. 0.0096;
Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 3 NYIFRAHDSVQTVIAKII 22
:| | | | | | | | | | | | | | | | | |
Db 554 SYSFRAHDSVQDLIADII 573
```

```
RESULT 7
US-09-290-049-15
; Sequence 15, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: FDC98-01p2A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 22
; TYPE: PRT
; ORGANISM: S. mutans
US-09-290-049-15
```

```
Query Match 63.6%; Score 68; DB 16; Length 22;
Best Local Similarity 70.0%; Pred. No. 0.00032;
Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 3 NYIFRAHDSVQTVIAKII 22
:| | | | | | | | | | | | | | | | | |
Db 3 SYSFRAHDSVQDLIRNII 22
```

```
RESULT 8
US-09-740-274-4
; Sequence 4, Application US/09740274
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
```

```
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-4
```

```
Query Match 63.6%; Score 68; DB 21; Length 1375;
Best Local Similarity 70.0%; Pred. No. 0.043;
Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 3 NYIFRAHDSVQTVIAKII 22
:| | | | | | | | | | | | | | | | | |
Db 580 SYSFRAHDSVQDLIRNII 599
```

```
RESULT 9
US-09-290-049-19
; Sequence 19, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: FDC98-01p2A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 19
; LENGTH: 22
; TYPE: PRT
; ORGANISM: S. sobrinus
US-09-290-049-19
```

```
Query Match 59.8%; Score 64; DB 16; Length 22;
Best Local Similarity 65.0%; Pred. No. 0.0016;
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
```

```
Qy 3 NYIFRAHDSVQTVIAKII 22
:| | | | | | | | | | | | | | | | | |
Db 3 SYSFRAHDSVQDIIRDII 22
```

```
RESULT 10
US-09-290-049-17
; Sequence 17, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
```

; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: FDC98-01p2A  
; CURRENT APPLICATION NUMBER: US/09/290,049  
; CURRENT FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081,550  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/115,142  
; EARLIER FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 17  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. downei  
US-09-290-049-17

Query Match 57.9%; Score 62; DB 16; Length 22;  
Best Local Similarity 65.0%; Pred. No. 0.0035;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 3 NYIFRAHDSVQTVIAKII 22  
:| | ||||| :| |  
Db 3 SYSFARHDSVQDLIRDI 22

RESULT 11  
US-09-290-049-2  
; Sequence 2, Application US/09290049  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: FDC98-01p2A  
; CURRENT APPLICATION NUMBER: US/09/290,049  
; CURRENT FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081,550  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/115,142  
; EARLIER FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: HDS peptide  
US-09-290-049-2

Query Match 50.9%; Score 54.5; DB 16; Length 20;  
Best Local Similarity 72.2%; Pred. No. 0.062;  
Matches 13; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

Qy 3 NYIFIR-AHDSEVQTVIA 19  
:| | | | | :| |  
Db 3 SYSFIRAHDSVQDLIA 20

RESULT 12  
US-09-499-203-2  
; Sequence 2, Application US/09499203  
; GENERAL INFORMATION:  
; APPLICANT: KOSSMANN, Jens  
; APPLICANT: WELSH, Thomas  
; APPLICANT: QUANZ, Martin  
; APPLICANT: KNUTH, Karola  
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase  
; FILE REFERENCE: 147-196P

; CURRENT APPLICATION NUMBER: US/09/499,203  
; CURRENT FILING DATE: 2000-02-08  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2057  
; TYPE: PRT  
; ORGANISM: Leuconostoc mesenteroides  
US-09-499-203-2

Query Match 49.5%; Score 53; DB 18; Length 2057;  
Best Local Similarity 55.6%; Pred. No. 27;  
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 3 NYIFRAHDSVQTVIAK 20  
:| | | | | :| |  
Db 759 NYSEVRAHDYDAQDPPIRK 776

RESULT 13  
US-60-191-637-21665  
; Sequence 21665, Application US/60191637  
; GENERAL INFORMATION:  
; APPLICANT: Venter, J. Craig  
; TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING  
; TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME, AND  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: CL000392  
; CURRENT APPLICATION NUMBER: US/60/191,637  
; CURRENT FILING DATE: 2000-03-23  
; NUMBER OF SEQ ID NOS: 42660  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 21665  
; LENGTH: 474  
; TYPE: PRT  
; ORGANISM: DROSOPHILA  
US-60-191-637-21665

Query Match 43.9%; Score 47; DB 24; Length 474;  
Best Local Similarity 47.1%; Pred. No. 52;  
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 4 YIFIRAHDSVQTVIAK 20  
:| | :| | :| | :| |  
Db 279 YVFVAHTSAGKTVAAE 295

RESULT 14  
US-60-191-681-17090  
; Sequence 17090, Application US/60191681  
; GENERAL INFORMATION:  
; APPLICANT: Li, Peter, W.D.  
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING ISOLATED DROSOPHILA PROTEINS AND  
; TITLE OF INVENTION: USES THEREOF.  
; FILE REFERENCE: CI000390  
; CURRENT APPLICATION NUMBER: US/60/191,681  
; CURRENT FILING DATE: 2000-03-23  
; NUMBER OF SEQ ID NOS: 30973  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 17090  
; LENGTH: 474  
; TYPE: PRT  
; ORGANISM: DROSOPHILA  
US-60-191-681-17090

Query Match 43.9%; Score 47; DB 24; Length 474;  
Best Local Similarity 47.1%; Pred. No. 52;  
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;



Qy 4 YIFIRAHDSVQTVIAK 20  
|:|: || | :||:|  
Db 279 YVFVAHTSAGKTVVAE 295

RESULT 15  
US-60-150-582-627  
; Sequence 627, Application US/60150582  
; GENERAL INFORMATION:  
; APPLICANT: Bonazzi, Vivien  
; TITLE OF INVENTION: ISOLATED KINASE PROTEINS, NUCLEIC ACID  
; TITLE OF INVENTION: MOLECULES ENCODING KINASE PROTEINS, AND USES THEREOF  
; FILE REFERENCE: CLO00084  
; CURRENT APPLICATION NUMBER: US/60/150.582  
; CURRENT FILING DATE: 1999-08-25  
; NUMBER OF SEQ ID NOS: 668  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 627  
; LENGTH: 1028  
; TYPE: PRT  
; ORGANISM: Drosophila  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(1028)  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-60-150-582-627

Query Match 43.98; Score 47; DB 24; Length 1028;  
Best Local Similarity 47.1%; Pred. No. 1.3e+02;  
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 4 YIFIRAHDSVQTVIAK 20  
|:|: || | :||:|  
Db 273 YVFVAHTSAGKTVVAE 289

Search completed: March 27, 2002, 14:20:29  
Job time: 1576 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:22:48 ; Search time 137.48 Seconds  
(without alignments)  
11.042 Million cell updates/sec

Title: US-09-290-049A-16

Perfect score: 107

Sequence: 1 MANYIFIRAHDSVQTVIAKII 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 332938 seqs, 6899538 residues

Total number of hits satisfying chosen parameters: 332938

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_New: \*  
1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep: \*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep: \*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep: \*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep: \*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep: \*  
6: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep1: \*  
7: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep: \*  
8: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107	100.0	22	6	US-09-562-328-29
2	107	100.0	22	6	US-09-290-049A-16
3	107	100.0	545	6	US-09-604-957-4
4	90	84.1	22	6	US-09-562-328-31
5	90	84.1	22	6	US-09-290-049A-18
6	86	80.4	523	6	US-09-604-957-5
7	68	63.6	22	6	US-09-562-328-28
8	68	63.6	22	6	US-09-290-049A-15
9	65	60.7	19	6	US-09-562-328-27
10	65	60.7	19	6	US-09-290-049A-2
11	65	60.7	23	6	US-09-562-328-44
12	64	59.8	22	6	US-09-562-328-32
13	64	59.8	22	6	US-09-290-049A-19
14	62	57.9	22	6	US-09-562-328-30
15	62	57.9	22	6	US-09-290-049A-17
16	55	51.4	584	6	US-09-604-957-6
17	54.5	50.9	20	6	US-09-562-328-26
18	50	46.7	535	6	US-09-604-957-7
19	50	46.7	1278	6	US-09-604-957-3
20	47	43.9	474	6	US-09-614-150-21600
21	46.5	43.5	866	6	US-09-614-150-21291
22	45	42.1	665	6	US-09-614-150-4278
23	44	41.1	575	6	US-09-708-427-10938
24	44	41.1	577	6	US-09-708-427-10937
25	44	41.1	583	6	US-09-708-427-10936

```

26 42 39.3 327 1 PCT-US02-03987-15771, A
27 42 39.3 327 7 US-10-032-585-7827, Ap
28 42 39.3 327 7 US-10-072-851-15771, A
29 42 39.3 336 6 US-09-614-150-3054, Ap
30 41.5 38.8 334 6 US-09-620-28541-311, Ap
31 41.5 38.8 354 1 PCT-US01-28541-2, Appli
32 40 37.4 147 7 US-10-015-127-10971, A
33 40 37.4 1286 8 US-60-337-358-618, App
34 39 36.4 182 5 US-09-901-484A-133, App
35 39 36.4 182 6 US-09-853-526-133, App
36 39 36.4 185 5 US-09-901-484A-136, App
37 39 36.4 185 6 US-09-853-526-136, App
38 39 36.4 228 5 US-09-901-484A-70, App
39 39 36.4 228 6 US-09-853-526-70, Appl
40 39 36.4 273 1 PCT-US02-03987-11100, A
41 39 36.4 273 6 US-09-815-242-11100, A
42 39 36.4 273 7 US-10-072-851-11100, A
43 39 36.4 300 5 US-09-901-484A-135, App
44 39 36.4 300 6 US-09-853-526-135, App
45 39 36.4 315 5 US-09-901-484A-134, App

```

#### ALIGNMENTS

RESULT 1

US-09-562-328-29

; Sequence 29, Application US/09562328

; GENERAL INFORMATION:

; APPLICANT: LEES, ANDREW

; APPLICANT: TAUBMAN, MARTIN A.

; APPLICANT: SMITH, DANIEL J.

; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES

; FILE REFERENCE: 04995.0046-01

; CURRENT APPLICATION NUMBER: US/09/562.328

; CURRENT FILING DATE: 2000-05-01

; PRIOR APPLICATION NUMBER: 09/288,965

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 44

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 29

; LENGTH: 22

; TYPE: PRT

; ORGANISM: Streptococcus sp.

US-09-562-328-29

Query Match 100.0%; Score 107; DB 6; Length 22;  
Best Local Similarity 100.0%; Pred. No. 1e-11;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MANYIFIRAHDSVQTVIAKII 22

Db 1 MANYIFIRAHDSVQTVIAKII 22

RESULT 2

US-09-290-049A-16

; Sequence 16, Application US/09290049A

; GENERAL INFORMATION:

; APPLICANT: SMITH, DANIEL J.

; APPLICANT: TAUBMAN, MARTIN A.

; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL

; FILE REFERENCE: 1564.1008-002

; CURRENT APPLICATION NUMBER: US/09/290.049A

; CURRENT FILING DATE: 1999-04-12

; PRIOR APPLICATION NUMBER: 60/081,550

; PRIOR FILING DATE: 1998-04-13

; PRIOR APPLICATION NUMBER: 60/115,142

; PRIOR FILING DATE: 1999-01-08

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 16  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. mutans  
US-09-290-049A-16

Query Match 100.0%; Score 107; DB 6; Length 22;  
Best Local Similarity 100.0%; Pred. No. 1e-11;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MANYIFIRAHDSVQTVIAKII 22  
|||:|||||  
Db 1 MANYIFIRAHDSVQTVIAKII 22

RESULT 3  
US-09-604-957-4  
; Sequence 4, Application US/09604957  
; GENERAL INFORMATION:  
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
; APPLICANT: DIJKHUIZEN, LUBBERT  
; APPLICANT: RAHAOUI, HAKIM  
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
; FILE REFERENCE: BO 43388  
; CURRENT APPLICATION NUMBER: US/09/604,957  
; CURRENT FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 00201871.1  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 545  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-604-957-4

Query Match 100.0%; Score 107; DB 6; Length 545;  
Best Local Similarity 100.0%; Pred. No. 6e-10;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MANYIFIRAHDSVQTVIAKII 22  
|||:|||||  
Db 154 MANYIFIRAHDSVQTVIAKII 175

RESULT 4  
US-09-562-328-31  
; Sequence 31, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995,0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 31  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-31

Query Match 84.1%; Score 90; DB 6; Length 22;  
Best Local Similarity 90.0%; Pred. No. 7.6e-09;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 NYIFIRAHDSVQTVIAKII 22  
|||:|||||  
Db 3 NYIFIRAHDSVQTVIAKII 22

RESULT 5  
US-09-290-049A-18  
; Sequence 18, Application US/09290049A  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CRIES  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 18  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. downei  
US-09-290-049A-18

Query Match 84.1%; Score 90; DB 6; Length 22;  
Best Local Similarity 90.0%; Pred. No. 7.6e-09;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 NYIFIRAHDSVQTVIAKII 22  
|||:|||||  
Db 3 NYIFIRAHDSVQTVIAKII 22

RESULT 6  
US-09-604-957-5  
; Sequence 5, Application US/09604957  
; GENERAL INFORMATION:  
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
; APPLICANT: DIJKHUIZEN, LUBBERT  
; APPLICANT: RAHAOUI, HAKIM  
; APPLICANT: LEER, ROBERT-JAN  
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
; FILE REFERENCE: BO 43388  
; CURRENT APPLICATION NUMBER: US/09/604,957  
; CURRENT FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 00201871.1  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 523  
; TYPE: PRT  
; ORGANISM: Leuconostoc mesenteroides  
US-09-604-957-5

Query Match 80.4%; Score 86; DB 6; Length 523;  
Best Local Similarity 80.0%; Pred. No. 2.1e-06;  
Matches 16; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 3 NYIFIRAHDSVQTVIAKII 22  
|||:|||||  
Db 148 NYIFIRAHDSVQTVIAKII 167

RESULT 7  
US-09-562-328-28

; Sequence 28, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 28  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-28

Query Match 63.6%; Score 68; DB 6; Length 22;  
Best Local Similarity 70.0%; Pred. No. 4.1e-05;  
Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
Qy 3 NYIFIRAHDSVQTVIAKII 22  
Db 3 SYSFIRAHDSVQDLIRNII 22  
:|||||:|:|

RESULT 8  
US-09-290-049A-15  
; Sequence 15, Application US/09290049A  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. mutans  
US-09-290-049A-15

Query Match 63.6%; Score 68; DB 6; Length 22;  
Best Local Similarity 70.0%; Pred. No. 4.1e-05;  
Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
Qy 3 NYIFIRAHDSVQTVIAKII 22  
Db 3 SYSFIRAHDSVQDLIRNII 22  
:|||||:|:|

RESULT 9  
US-09-562-328-27  
; Sequence 27, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01

; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 27  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-27

Query Match 60.7%; Score 65; DB 6; Length 19;  
Best Local Similarity 76.5%; Pred. No. 0.00011;  
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
Qy 3 NYIFIRAHDSVQTVIA 19  
Db 3 SYSFIRAHDSVQDLIA 19  
:|||||:|:|

RESULT 10  
US-09-290-049A-2  
; Sequence 2, Application US/09290049A  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: HDS peptide  
US-09-290-049A-2

Query Match 60.7%; Score 65; DB 6; Length 19;  
Best Local Similarity 76.5%; Pred. No. 0.00011;  
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
Qy 3 NYIFIRAHDSVQTVIA 19  
Db 3 SYSFIRAHDSVQDLIA 19  
:|||||:|:|

RESULT 11  
US-09-562-328-47  
; Sequence 44, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 44  
; LENGTH: 23  
; TYPE: PRT

;  
US-09-562-328-44

Query Match 60.7%; Score 65; DB 6; Length 23;  
Best Local Similarity 76.5%; Pred. No. 0.00014;  
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 NYIFIRAHDSVQTVIA 19  
Db 7 SYSFARAHDSVQDLIA 23

RESULT 12  
US-09-562-328-32  
; Sequence 32, Application US/09562328  
; GENERAL INFORMATION:

APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 32  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-32

Query Match 59.8%; Score 64; DB 6; Length 22;  
Best Local Similarity 65.0%; Pred. No. 0.0002;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 NYIFIRAHDSVQTVIAKII 22  
Db 3 SYSFARAHDSVQDLIRII 22

RESULT 13  
US-09-290-049A-19  
; Sequence 19, Application US/09290049A  
; GENERAL INFORMATION:

APPLICANT: Smith, Daniel J.  
APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 19  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. sobrinus  
US-09-290-049A-19

Query Match 59.8%; Score 64; DB 6; Length 22;  
Best Local Similarity 65.0%; Pred. No. 0.0002;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 NYIFIRAHDSVQTVIAKII 22

Db 3 SYSFARAHDSVQDLIRII 22

RESULT 14  
US-09-562-328-30  
; Sequence 30, Application US/09562328  
; GENERAL INFORMATION:

APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 30  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-30

Query Match 57.9%; Score 62; DB 6; Length 22;  
Best Local Similarity 65.0%; Pred. No. 0.00043;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 NYIFIRAHDSVQTVIAKII 22  
Db 3 SYSFARAHDSVQDLIRII 22

RESULT 15  
US-09-290-049A-17  
; Sequence 17, Application US/09290049A  
; GENERAL INFORMATION:

APPLICANT: Smith, Daniel J.  
APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. downei  
US-09-290-049A-17

Query Match 57.9%; Score 62; DB 6; Length 22;  
Best Local Similarity 65.0%; Pred. No. 0.00043;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 NYIFIRAHDSVQTVIAKII 22  
Db 3 SYSFARAHDSVQDLIRII 22

Search completed: March 27, 2002, 14:22:49  
Job time: 1696 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:01:22 ; Search time 102.51 seconds  
(without alignments)  
16.348 Million cell updates/sec

Title: US-09-290-049A-16  
Perfect score: 107  
Sequence: 1 MANYIFIRAHDSVQTVIAKII 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	1431	2 A45866	dextranucrase (EC
2	90	84.1	1365	2 A41483	glucosyltransferas
3	89	83.2	1577	2 T30858	glucosyltransferas
4	88	82.2	1449	2 T30857	glucosyltransferas
5	88	82.2	1449	2 T30552	glucosyltransferas
6	87	81.3	1508	2 T31098	probable dextranu
7	87	81.3	1599	2 S22737	glucosyltransferas
8	85	79.4	1290	2 J05473	dextranucrase (EC
9	78	72.9	1518	2 A44811	glucosyltransferas
10	72	67.3	1475	2 B31135	gtfB protein precu
11	68	63.6	1375	2 J70345	dextranucrase (EC
12	64	59.8	1592	2 A38175	glucosyltransferas
13	45	42.1	267	2 B84213	hypothetical prote
14	44.5	41.6	593	2 S51946	pyruvate kinase (E
15	44	41.1	132	2 D82080	mutator Mutr prote
16	44	41.1	431	2 S49821	PRL2 protein - Ara
17	44	41.1	583	2 T04531	nine-cis-epoxycaro
18	44	41.1	1213	2 T41378	probable helicase
19	43	40.2	288	2 T26303	hypothetical prote
20	43	40.2	933	2 G70166	probable zinc prot
21	42	39.3	236	2 E82593	phosphate regulon
22	42	39.3	1153	2 T31080	nitric-oxide synth
23	41.5	38.8	354	2 A70255	immunogenic protei
24	41.5	38.8	493	2 T10054	pyruvate kinase (E
25	41.5	38.8	583	2 T10051	pyruvate kinase (E
26	41	38.3	76	2 J01633	HCRF3 protein - hu
27	41	38.3	140	2 J70515	probable lipoprote
28	41	38.3	368	2 T40115	uv excision repair
29	41	38.3	411	2 A34526	ORF1 protein - Org

30 41 38.3 411 2 T10373 hypothetical prote  
31 41 38.3 602 2 A71256 aspartate--trna li  
32 41 38.3 725 2 T17732 helicase-like prot  
33 41 38.3 768 2 S52684 probable membrane  
34 41 38.3 1165 2 D72496 hypothetical prote  
35 41 38.3 1771 2 S76851 hypothetical prote  
36 40.5 37.9 217 1 H69797 rhamnogalacturonan  
37 40 37.4 344 1 GRYCS7 sulfate transport  
38 40 37.4 354 2 S65687 (A+T)-stretch-bind  
39 40 37.4 417 2 S25627 glucose-1-phosphat  
40 40 37.4 554 2 C69161 sensory transducti  
41 40 37.4 557 2 S73434 aspartate--trna li  
42 40 37.4 586 2 B84271 glutamyl-trna synt  
43 40 37.4 654 1 BVBYPI CBP1 protein - yea  
44 40 37.4 984 2 T50309 hypothetical WD-re  
45 40 37.4 1018 1 GNWXG7 genome polyprotein

ALIGNMENTS

RESULT 1  
A45866  
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans  
C:Species: Streptococcus mutans  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: A45866  
R:Honda, O.; Kato, C.; Kuramitsu, H. K.  
J. Gen. Microbiol. 136, 2099-2105, 1990  
A:Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the gluco  
A:Reference number: A45866; MUID:91100958  
A:Accession: A45866  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1431 <HON>  
A:Cross-references: GB:M29296  
C:Superfamily: cpl repeat homology  
C:Keywords: glycosyltransferase; hexosyltransferase  
F:181-201/Domain: cpl repeat homology <CP1>  
F:1127-1146/Domain: cpl repeat homology <CP2>  
F:1192-1211/Domain: cpl repeat homology <CP3>  
F:1257-1276/Domain: cpl repeat homology <CP4>  
F:1277-1297/Domain: cpl repeat homology <CP5>  
F:1321-1340/Domain: cpl repeat homology <CP6>  
F:1341-1361/Domain: cpl repeat homology <CP6>  
F:1385-1404/Domain: cpl repeat homology <CP7>

Query Match 100.0%; Score 107; DB 2; Length 1431;  
Best Local Similarity 100.0%; Pred. No. 1.6e-09;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MANYIFIRAHDSVQTVIAKII 22  
DB 574 MANYIFIRAHDSVQTVIAKII 595

RESULT 2  
A41483  
glucosyltransferase (EC 2.4.1.-) gtfS precursor - Streptococcus sobrinus  
C:Species: Streptococcus sobrinus  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 15-Oct-1999  
C:Accession: A41483  
R:Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.  
Infect. Immun. 58, 2452-2458, 1990  
A:Title: Analysis of the Streptococcus downei gtfS gene, which specifies a glucosyltr  
A:Reference number: A41483; MUID:90316665  
A:Accession: A41483  
A:Molecule type: DNA  
A:Residues: 1-1365 <GIL>  
A:Cross-references: GB:M30943; NID:gl53652; PIDN:AAA26898.1; PID:gl53653  
C:Genetics:  
A:Gene: gtfS

C:Superfamily: cpl repeat homology  
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 84.1%; Score 90; DB 2; Length 1365;  
Best Local Similarity 90.0%; Pred. No. 1.1e-06;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NYIFIRAHDSVQTVIAKII 22  
|||:|||||:|||||:|||||

Db 539 NYVFIKRAHDSVQTVIAKII 558

## RESULT 3

T30858

glucosyltransferase - Streptococcus salivarius

C:Species: Streptococcus salivarius

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999

C:Accession: T30858

R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.

Infect. Immun. 63, 609-621, 1995

A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pri

A:Reference number: Z20909; MUID:95122197

A:Accession: T30858

A&gt;Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-1577 &lt;SIM&gt;

A:Cross-references: EMBL:L35928; NID:g662380; PID:g662381; PIDN:AAC41413.1

C:Genetics:

A:Gene: gtfm

Query Match 83.2%; Score 89; DB 2; Length 1577;  
Best Local Similarity 81.0%; Pred. No. 1.9e-06;  
Matches 17; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 ANYFIRAHDSVQTVIAKII 22  
||||:|||||:|||||:|||||

Db 660 ANYFVRAHDSVQVLANII 680

## RESULT 4

T30857

glucosyltransferase - Streptococcus salivarius

C:Species: Streptococcus salivarius

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999

C:Accession: T30857

R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.

Infect. Immun. 63, 609-621, 1995

A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pri

A:Reference number: Z20909; MUID:95122197

A:Accession: T30857

A&gt;Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-1449 &lt;SIM&gt;

A:Cross-references: EMBL:L35495; NID:g662378; PID:g662379; PIDN:AAC41412.1

C:Genetics:

A:Gene: gtfL

Query Match 82.2%; Score 88; DB 2; Length 1449;  
Best Local Similarity 72.7%; Pred. No. 2.5e-06;  
Matches 16; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 MANYIFIRAHDSVQTVIAKII 22  
|||||:|||||:|||||:|||||

Db 607 MANYAFVRAHDSVQSIIGQII 628

## RESULT 5

T30552

glucosyltransferase N - Streptococcus salivarius (fragment)

C:Species: Streptococcus salivarius

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30552

R:Jaife, R.I.

submitted to the EMBL Data Library, February 1998

A:Description: Streptococcus salivarius V1477 gtfN.

A:Reference number: Z20854

A:Accession: T30552

A&gt;Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-1449 &lt;JAF&gt;

A:Cross-references: EMBL:AF049609; NID:g2935545; PID:g2935546; PIDN:AAC05156.1

C:Genetics:

A:Gene: gtfN

Query Match 82.2%; Score 88; DB 2; Length 1449;  
Best Local Similarity 72.7%; Pred. No. 2.5e-06;  
Matches 16; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 MANYIFIRAHDSVQTVIAKII 22  
|||||:|||||:|||||:|||||

Db 607 MANYAFVRAHDSVQSIIGQII 628

## RESULT 6

T31098

probable dextranucrase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides

C:Species: Leuconostoc mesenteroides

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000

C:Accession: T31098

R:Monchois, V.; Renaud-Simeon, M.; Monsan, P.; Willemot, R.M.

FEMS Microbiol. Lett. 159, 307-315, 1998

A:Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase

A:Reference number: Z20981; MUID:98164374

A:Accession: T31098

A&gt;Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-1508 &lt;MON&gt;

A:Cross-references: EMBL:AF030129; NID:g2766611; PID:g2766612; PIDN:AAB95453.1

A:Experimental source: strain NRRL B-1299

C:Genetics:

A:Gene: dsrB

C:Function:

A:Description: produces dextran composed only of alpha(1-6) glucosidic bonds

C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 81.3%; Score 87; DB 2; Length 1508;  
Best Local Similarity 85.0%; Pred. No. 3.9e-06;  
Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 NYIFIRAHDSVQTVIAKII 22  
|||:|||||:|||||:|||||

Db 636 NYSFVRAHDSVQTVIAQII 655

## RESULT 7

S22737

glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius

C:Species: Streptococcus salivarius

C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 21-Jan-2000

C:Accession: S22737; S28810; B44811; S22727

R:Jacques, N.

submitted to the EMBL Data Library, March 1992

A:Reference number: S22736

A:Accession: S22737

A:Molecule type: DNA

A:Residues: 1-1599 &lt;JAC&gt;

A:Cross-references: EMBL:Z11872; NID:g47530; PIDN:CAA77898.1; PID:g47531

A:Experimental source: ATCC 25975

R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.

J. Gen. Microbiol. 137, 2577-2593, 1991

A:Title: Molecular characterization of a cluster of at least two glucosyltransferase

A;Reference number: A44811; MUID:92148377

A;Accession: S28810

A;Molecule type: DNA

A;Residues: 1-51 <GIF>

A;Cross-references: EMBL:Z11873

C;Genetics:

A;Gene: gtfk

C;Superfamily: cpl repeat homology

C;Keywords: glycosyltransferase; hexosyltransferase

F;1456-1475/Domain: cpl repeat homology <CPR>

Query Match 81.3%; Score 87; DB 2; Length 1599;

Best Local Similarity 81.0%; Pred. No. 4.2e-06;

Matches 17; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 ANYFIRAHDSVQTVIAKII 22

I:|||||:|||||:|||||:|||||

Db 573 ATYLFVRAHDSVQTVIADII 593

LT 8

73

dextranucrase (EC 2.4.1.5) - Leuconostoc mesenteroides

C;Species: Leuconostoc mesenteroides

C;Date: 07-Jul-1997 #sequence\_revision 29-Aug-1997 #text\_change 29-Aug-1997

C;Accession: JC5473

R;Monchois, V.; Willemot, R.M.; Remaud-Simeon, M.; Croux, C.; Monsan, P.

Gene 182, 23-32, 1996

A;Title: Cloning and sequencing of a gene coding for a novel dextranucrase from Leucon

A;Reference number: JC5473; MUID:97136686

A;Accession: JC5473

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-1290 <NON>

A;Cross-references: GB:U38181

C;Comment: This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose ont

C;Genetics:

A;Gene: dsrA

C;Keywords: glycosyltransferase; hexosyltransferase

F;78-870/Domain: catalytic #status predicted <CAT>

F;922-1290/Domain: glucan-binding #status predicted <GCB>

Query Match 79.4%; Score 85; DB 2; Length 1290;

Best Local Similarity 85.0%; Pred. No. 7.1e-06;

Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 NYIFIRAHDSVQTVIAKII 22

I:|||||:|||||:|||||:|||||

389 NYSFIRAHDSVQTVIADII 408

RESULT 9

A44811

glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius

C;Species: Streptococcus salivarius

C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 15-Oct-1999

C;Accession: A44811; S22726; S28809

R;Giffard, P.N.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.

J. Gen. Microbiol. 137, 2577-2593, 1991

A;Title: Molecular characterization of a cluster of at least two glucosyltransferase gen

A;Reference number: A44811; MUID:92148377

A;Accession: A44811

A;Molecule type: DNA

A;Residues: 1-1518 <GIF>

A;Cross-references: EMBL:Z11873; NID:g47526; PIDN:CAA77900.1; PID:g47527

A;Note: sequence extracted from NCBI backbone (NCBI:81050, NCBIP:81052)

C;Genetics:

A;Gene: gtfJ

C;Superfamily: cpl repeat homology

C;Keywords: glycosyltransferase; hexosyltransferase

F;1307-1326/Domain: cpl repeat homology <CP4>

Query Match 72.9%; Score 78; DB 2; Length 1518;

Best Local Similarity 70.0%; Pred. No. 0.00013;

Matches 14; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 NYIFIRAHDSVQTVIAKII 22

I:|||||:|||||:|||||:|||||

Db 604 NYVFIARHNNVQDIIEII 623

RESULT 10

B33135

gtfB protein precursor - Streptococcus mutans

C;Species: Streptococcus mutans

C;Date: 23-Oct-1990 #sequence\_revision 23-Oct-1990 #text\_change 15-Oct-1999

C;Accession: B33135; A33128

R;Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

J. Bacteriol. 169, 4263-4270, 1987

A;Title: Sequence analysis of the gtfB gene from Streptococcus mutans.

A;Reference number: A33135; MUID:87308013

A;Accession: B33135

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1475 <SHI>

A;Cross-references: GB:M17361; NID:g153639; PIDN:AAA88588.1; PID:g153640

R;Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

submitted to the Protein Sequence Database, September 1990

A;Reference number: A33128

A;Accession: A33128

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: DNA

A;Residues: 1-171,173-641,'N' 643-1475 <SH2>

A;Experimental source: strain GS-5

C;Superfamily: cpl repeat homology

F;1096-1115/Domain: cpl repeat homology <CP1>

F;1224-1243/Domain: cpl repeat homology <CP2>

F;1289-1308/Domain: cpl repeat homology <CP3>

F;1354-1373/Domain: cpl repeat homology <CP4>

F;1419-1438/Domain: cpl repeat homology <CP5>

Query Match 67.3%; Score 72; DB 2; Length 1475;

Best Local Similarity 75.0%; Pred. No. 0.0013;

Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 NYIFIRAHDSVQTVIAKII 22

I:|||||:|||||:|||||:|||||

Db 554 SYSFIRAHDSVQDLIADII 573

RESULT 11

JT0345

dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)

N;Alternate names: sucrose 6-glucosyltransferase

C;Species: Streptococcus mutans

C;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 24-Sep-1999

C;Accession: JT0345; C33135

R;Ueda, S.; Shiroza, T.; Kuramitsu, H.K.

Gene 69, 101-109, 1988

A;Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.

A;Reference number: JT0345; MUID:89137980

A;Accession: JT0345

A;Molecule type: DNA

A;Residues: 1-1375 <UED>

A;Experimental source: GS-5

R;Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

J. Bacteriol. 169, 4263-4270, 1987

A;Title: Sequence analysis of the gtfB gene from Streptococcus mutans.

A;Reference number: A33135; MUID:87308013

A;Accession: C33135

A;Status: preliminary

A;Molecule type: DNA

A;Molecule  
A;Residues  
A;Cross-re

A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VC2392  
A:Map position: 1  
C:Superfamily: mutator mutT; mutT domain homology

Query Match 41.1%; Score 44; DB 2; Length 132;  
Best Local Similarity 50.0%; Pred. No. 4.3;  
Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MANYIFIRAHDSVOTVIK 20  
:|||||:|:|:|:  
Db 111 LANYRPEANDPVVKQVIAQ 130

Search completed: March 27, 2002, 14:01:23  
Job time: 485 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:27:06 ; Search time 53.4 seconds  
(without alignments)  
15.105 Million cell updates/sec

Title: US-09-290-049a-16  
Perfect score: 107  
Sequence: 1 MANYIFIRAHDSVQTVIAKII 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	107	100.0	1462	1	GTFD_STRMU
2	90	84.1	1365	1	GTFS_STRDO
3	68	63.6	1375	1	GTFC_STRMU
4	67	62.6	1476	1	GTFB_STRMU
5	64	59.8	1592	1	GTFL_STRDO
6	62	57.9	1597	1	GTFL_STRDO
7	44.5	41.6	593	1	KPYA_TOBAC
8	44	41.1	479	1	PRL2_ARATH
9	42	39.3	205	1	ADEN_ADEG8
10	42	39.3	236	1	PHOU_XYLFA
11	42	39.3	336	1	RA51_DROME
12	42	39.3	1153	1	NOS_LYMSI
13	41.5	38.8	583	1	KPYA_RICCO
14	41	38.3	411	1	VP48_NPVOP
15	41	38.3	602	1	SYD_TREPA
16	40	37.4	344	1	CYSA_SYNP7
17	40	37.4	417	1	AGP_PRORE
18	40	37.4	557	1	SYD_MYCPN
19	40	37.4	654	1	CBP1_YEAST
20	40	37.4	1018	1	VGNM_BPMV
21	40	37.4	1246	1	SKIW_HUMAN
22	40	37.4	1287	1	SKI2_YEAST
23	39.5	36.9	231	1	BID0_VIRCH
24	39.5	36.9	332	1	COBS_PSEDE
25	39.5	36.9	1082	1	RPB0_ROTPE
26	39	36.4	238	1	Y381_TREPA
27	39	36.4	272	1	RL2_HAEIN
28	39	36.4	283	1	LEG1_HAECO
29	39	36.4	297	1	LE33_CAEEL
30	39	36.4	350	1	VDD2_DROME
31	39	36.4	351	1	RFB8_XANCA
32	39	36.4	380	1	ADH_MALDO
33	39	36.4	433	1	TCR_STAUA

34	39	36.4	525	1	COX1_CAEEL
35	39	36.4	548	1	LIP2_CANRU
36	39	36.4	549	1	LIP3_CANRU
37	39	36.4	549	1	LIP4_CANRU
38	39	36.4	586	1	SYD_BORBU
39	39	36.4	606	1	SP2_HUMAN
40	39	36.4	3678	1	DMD_MOUSE
41	38	35.5	135	1	Y548_BUCAI
42	38	35.5	187	1	VGG_BPPHK
43	38	35.5	317	1	YG00_HAEIN
44	38	35.5	367	1	NUIC_MESVI
45	38	35.5	505	1	SPKD_SYNY3

ALIGNMENTS

RESULT 1	GTFD_STRMU	STANDARD:	PRT: 1462 AA.
ID	AC	P49331: O69383; O69386; O69389; O69392; O69398;	
DT	01-FEB-1996	(Rel. 33, Created)	
DT	15-JUL-1999	(Rel. 38, Last sequence update)	
DT	15-JUL-1999	(Rel. 38, Last annotation update)	
DE	GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)		
DE	(SUCROSE 6-GLUCOSYLTRANSFERASE).		
GN	GTFD.		
OS	Streptococcus mutans.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;		
OC	Streptococcus.		
OX	NCBI_TaxID=1309;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=GS-5;		
RC	MEDLINE=91100958; PubMed=2148600;		
RX	Honda O., Kato C., Kuramitsu H.K.;		
RA	"Nucleotide sequence of the Streptococcus mutans gtfd gene encoding		
RT	the glucosyltransferase-S enzyme."		
RL	J. Gen. Microbiol. 136:2099-2105(1990).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MT2239, MT4245, MT4251, MT4467, AND MT8148;		
RC	MEDLINE=98231643; PubMed=9570124;		
RX	Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,		
RA	Kimura S., Hada S.;		
RT	"Molecular analyses of glucosyltransferase genes among strains of		
RL	Streptococcus mutans."		
RL	FEMS Microbiol. Lett. 161:331-336(1998).		
CC	FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT		
CC	TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE		
CC	OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE		
CC	AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.		
CC	CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -		
CC	D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).		
CC	SUBCELLULAR LOCATION: SECRETED.		
CC	DISEASE: DENTAL CARIES.		
CC	MISCELLANEOUS: GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA		
CC	1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES). GTF-S SYNTHESIZES		
CC	WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSYL). GTF-SI SYNTHESIZES BOTH		
CC	FORMS OF GLUCANS.		
CC	SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-		
CC	BINDING PROTEIN FROM S. MUTANS.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; M29296; AAA26895.1; -		
DR	EMBL; D88653; BAA26103.1; -		

```
DR EMBL; D88656; BAA26107.1; -.
DR EMBL; D88659; BAA26111.1; -.
DR EMBL; D88662; BAA26115.1; -.
DR EMBL; D89979; BAA26121.1; -.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 11.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1 ?
FT CHAIN ? 1462
FT DOMAIN 1232 1423
FT REPEAT 1232 1295
FT REPEAT 1296 1359
FT REPEAT 1360 1423
FT VARIANT 58 68
FT VARIANT 68 81
FT VARIANT 81 81
FT VARIANT 113 113
FT VARIANT 113 122
FT VARIANT 132 132
FT VARIANT 135 135
FT VARIANT 202 202
FT VARIANT 255 255
FT VARIANT 275 275
FT VARIANT 288 288
FT VARIANT 301 301
FT VARIANT 313 313
FT VARIANT 317 317
FT VARIANT 328 328
FT VARIANT 350 350
FT VARIANT 628 633
FT VARIANT 688 688
FT VARIANT 726 732
FT VARIANT 726 730
FT VARIANT 964 964
FT VARIANT 1019 1019
FT VARIANT 1059 1060
FT VARIANT 1060 1060
FT VARIANT 1080 1080
FT VARIANT 1142 1142
FT VARIANT 1198 1198
FT VARIANT 1220 1220
FT VARIANT 1280 1280
FT VARIANT 1282 1282
FT VARIANT 1290 1290
FT VARIANT 1311 1311
FT VARIANT 1403 1403
FT VARIANT 1425 1425
FT VARIANT 1449 1449
FT CONFLICT 1428 1462
FT VYR (IN REF. 1).
SQ SEQUENCE 1462 AA; 163512 MW; 5C6541F0DCB0DF00 CRC64;

GTFS_STRDO STANDARD; PRT; 1365 AA.
AC P29336;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTF-S.
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1317;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MFE28;
RX MEDLINE=90316665; PubMed=2142479;
RA "Analysis of the Streptococcus downei gtfS gene, which specifies a
RT glucosyltransferase that synthesizes soluble glucans.";
RL Infect. Immun. 58:2452-2458(1990).
CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -!- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).
CC -!- ENZYME REGULATION: GLUCAN SYNTHESIS BY GTF-S IS INDEPENDENT OF
CC PRIMER GLUCAN UNLIKE GTF-I.
CC -!- DISEASE: DENTAL CARIES.
CC -!- MISCELLANEOUS: GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA
CC 1,6-GLUCOSE).
CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S.MUTANS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M30943; AAA26898.1; -.
DR PIR; A41483; A41483.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 10.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1 36
FT CHAIN 37 1365
FT DOMAIN 37 1050
FT DOMAIN 1083 1365
FT DOMAIN 1083 1365
FT REPEAT 1083 1131
FT REPEAT 1150 1199
FT REPEAT 1225 1274
FT REPEAT 1289 1339
FT REPEAT 1353 1365
FT SEQUENCE 1365 AA; 151590 MW; 167296B5A2E8C476 CRC64;
```

Query Match 84.1%; Score 90; DB 1; Length 1365;  
Best Local Similarity 90.0%; Pred. No. 2.9e-07;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NYIFIRAHDSVQTVIAKII 22  
|||||  
DB 539 NYIFIRAHDSVQTVIAKII 558

Query Match 100.0%; Score 107; DB 1; Length 1462;  
Best Local Similarity 100.0%; Pred. No. 3.8e-10;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MANYIFIRAHDSVQTVIAKII 22  
|||||  
DB 574 MANYIFIRAHDSVQTVIAKII 595





```
RESULT 3
ID GTFC_STRMU STANDARD; PRT; 1375 AA.
AC P13470; 205427;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI PRECURSOR (EC 2.4.1.5) (GTF-SI)
DE (DEXTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTFC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=89137980; PubMed=2976010;
RA Ueda S., Shiroza T., Kuramitsu H.K.;
RT "Sequence analysis of the gtfc gene from Streptococcus mutans GS-5.";
PB Gene 69:101-109(1988).
[2]
SEQUENCE OF 1-349 FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfb gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) =
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DISEASE: DENTAL CARIES.
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S. MUTANS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M22054; AAA88592.1; -
EMBL; M17361; AAA88589.1; -
PIR; JT0345; JT0345.
DR InterPro; IPR003318; Glyco_hydro_70.
DR InterPro; IPR002479; CW_binding.
DR Pfam; PF01473; CW_binding_1; 7.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1 34
FT CHAIN 35 1375
FT DOMAIN 35 1050 GLUCOSYLTRANSFERASE-SI.
FT DOMAIN 1126 1375 CATALYTIC (APPROXIMATE).
FT DOMAIN 1126 1375 GLUCAN-BINDING (APPROXIMATE).
FT REPEAT 1126 1159 2.4 A, 1 C AND 1 AC REPEATS.
FT REPEAT 1169 1200 A REPEAT.
FT REPEAT 1227 1238 C REPEAT.
FT REPEAT 1253 1303 C REPEAT.
FT REPEAT 1318 1330 AC REPEAT.
FT REPEAT 1375 AA; 153022 MW; D4B80CBEE0AACE13 CRC64;
SQ SEQUENCE 1375 AA; 153022 MW; D4B80CBEE0AACE13 CRC64;

Query Match 63.6%; Score 68; DB 1; Length 1375;
Best Local Similarity 70.0%; Pred. No. 0.0017;
Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 3 NYIFRAHDSVQTVIAKII 22
Db 580 SYSFIRAHDSVQDLIRNII 599

RESULT 4
ID GTFB_STRMU STANDARD; PRT; 1476 AA.
AC P08987; 069381; 069384; 069390; 069396;
DT 01-NOV-1988 (Rel. 09, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTFB.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfb gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
[2]
SEQUENCE FROM N.A.
RC STRAIN=MT4239, MT4245, MT4251, MT4467, AND MT8148;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) =
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DISEASE: DENTAL CARIES.
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S. MUTANS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M17361; AAA88588.1; -
EMBL; D88651; BAA26101.1; -
EMBL; D88654; BAA26105.1; -
EMBL; D88657; BAA26109.1; -
EMBL; D88660; BAA26113.1; -
EMBL; D89577; BAA26119.1; -
PIR; B33135; B33135.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 13.
DR Pfam; PF02324; Glyco_hydro_70; 1.
```











FT SEQUENCE 583 AA; 64093 MW; 5DFD84F778A753C2 CRC64;  
SQ

Query Match 38.8%; Score 41.5; DB 1; Length 583;  
Best Local Similarity 47.6%; Pred. No. 23;  
Matches 10; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

OY 1 MANYIFIRAHSEVQTVIAKI 21  
: ||| | : ||| |  
DB 313 LKSYTAARSDSDI-AVIKI 332

## RESULT 14

ID VP48\_NPVOP STANDARD; PRT; 411 AA.  
AC P24651;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
P48 PROTEIN.

Orygia pseudotsugata multicapsid polyhedrosis virus (OpNPV).  
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
OC Nucleopolyhedrovirus.  
OX NCBI\_TaxID=164623;  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE=90188300; PubMed=2179466;  
RA Russell R.L.Q., Rohmann G.F.;  
RT "The p6.5 gene region of a nuclear polyhedrosis virus of Orygia  
pseudotsugata: DNA sequence and transcriptional analysis of four late  
genes";  
RL J. Gen. Virol. 71:551-560(1990).  
RN [2]

RP SEQUENCE FROM N.A.  
RX MEDLINE=90232722; PubMed=2184573;  
RA Mueller R., Pearson M., Russell R.L.Q., Rohmann G.F.;  
RT "A capsid-associated protein of the multicapsid nuclear polyhedrosis  
virus of Orygia pseudotsugata: genetic location, sequence,  
transcriptional mapping, and immunocytochemical characterization.";  
RL Virology 176:133-144(1990).  
RN [3]

RP SEQUENCE FROM N.A.  
RX MEDLINE=97271300; PubMed=9126251;  
RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,  
RA Rohmann G.F.;  
RT "The sequence of the Orygia pseudotsugata multinucleocapsid nuclear  
polyhedrosis virus genome";  
RL Virology 229:381-399(1997).

This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
-----  
DR EMBL; D13959; BAA03060.1; -;  
DR EMBL; U75930; AAC59103.1; -;  
DR PIR; A34526; A34526.  
KW Late protein.  
SQ SEQUENCE 411 AA; 47880 MW; BCB86A65EE30CD01 CRC64;

Query Match 38.3%; Score 41; DB 1; Length 411;  
Best Local Similarity 38.1%; Pred. No. 19;  
Matches 8; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

OY 1 MANYIFIRAHSEVQTVIAKI 21  
: ||| | : ||| |  
DB 342 VCRYIFKRYSDKDVAVVVEKL 362

## RESULT 15

ID SYD\_TREPA STANDARD; PRT; 602 AA.  
AC SYD\_TREPA;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE ASPARTYL-TRNA SYNTHETASE (EC 6.1.1.12) (ASPARTATE--TRNA LIGASE)  
DE (ASPRS).  
GN ASPs OR TP0985.  
OS Treponema pallidum.

OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.

OX NCBI\_TaxID=160;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NICHOLS;

RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,

RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,

RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,

RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Uterback T.,

RA McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,

RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,

RA Venter J.C.;

RT "Complete genome sequence of Treponema pallidum, the syphilis

RT Spirochete";

RL Science 281:375-388(1998).

CC -!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + TRNA(ASP) = AMP +

CC PYROPHOSPHATE + L-ASPARTYL-TRNA(ASP).

CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.

CC -!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
-----

DR EMBL; AE001266; AAC65942.1; -;

DR TIGR; TP0985; -;

DR InterPro; IPR002106; AA\_trna\_ligase\_II.

DR InterPro; IPR002309; trna-synt\_2.

DR InterPro; IPR002312; trna-synt\_2.

DR Pfam; PF00152; trna-synt\_2; 1.

DR Pfam; PF01336; trna-anti; 1.

DR PRINTS; PR01042; TRNASYNTHASP.

DR PROSITE; PS00179; AA\_TRNA\_LIGASE\_II\_1; 1.

DR PROSITE; PS00339; AA\_TRNA\_LIGASE\_II\_2; FALSE\_NEG.

KW Aminoacyl-TRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

KW Complete proteome.

SQ SEQUENCE 602 AA; 68626 MW; A51D0B17DE4FC5BC CRC64;

Query Match 38.3%; Score 41; DB 1; Length 602;

Best Local Similarity 50.0%; Pred. No. 29;

Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 7 IRAHSEVQTVIAKII 22

||| | : ||| |

DB 505 IRIHDTQKRFKIV 520

Search completed: March 27, 2002, 14:27:07

Job time: 1649 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:27:05 ; Search time 53.4 seconds  
(without alignments)  
15.105 Million cell updates/sec

Title: US-09-290-049A-15

Perfect score: 110

Sequence: 1 VPSYSFIRAHSEVDLIRNII 22

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

1 number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	110	100.0	1375	1	GTFC_STRMU
2	105	95.5	1476	1	GTFC_STRMU
3	100	90.9	1597	1	GTFC_STRMU
4	98	89.1	1592	1	GTFC_STRMU
5	76	69.1	1365	1	GTFC_STRMU
6	68	61.8	1462	1	GTFC_STRMU
7	48	43.6	540	1	ANPC_HUMAN
8	46	41.8	196	1	MAD2_YEAST
9	45	40.9	1835	1	DUR1_YEAST
10	45	40.9	6359	1	BACC_BACLI
11	44	40.0	282	1	YIAJ_ECOLI
12	44	40.0	344	1	M12D_BACSU
13	44	40.0	587	1	YJN4_YEAST
14	44	40.0	1068	1	P11A_BOVIN
15	44	40.0	1068	1	P11A_HUMAN
16	44	40.0	1068	1	P11A_MOUSE
17	43	39.1	214	1	COMA_BACSU
18	43	39.1	691	1	Y104_YEAST
19	43	39.1	2541	1	TALI_HUMAN
20	42	38.2	225	1	SRD1_YEAST
21	42	38.2	584	1	SYD_BUCAP
22	42	38.2	932	1	HRL_SCHPO
23	42	38.2	4969	1	RYNC_RABIT
24	41.5	37.7	218	1	GPH_RHOSH
25	41.5	37.7	713	1	DC12_ECOLI
26	41	37.3	183	1	YS4L_CAEEL
27	41	37.3	242	1	SGAE_MYCPN
28	41	37.3	344	1	MBHS_AZCH
29	41	37.3	537	1	ANPC_BOVIN
30	41	37.3	659	1	RNB_HAEIN
31	41	37.3	875	1	ENV_BIV06
32	41	37.3	904	1	ENV_BIV27
33	41	37.3	1006	1	K125_TOBAC

34 40.5 36.8 63 1 V7K\_BYDVP  
35 40.5 36.8 183 1 RETB\_BOVIN  
36 40.5 36.8 201 1 RETB\_PIG  
37 40 36.4 96 1 IBBA\_PEA  
38 40 36.4 114 1 IBB2\_PEA  
39 40 36.4 205 1 ADEN\_ADEG8  
40 40 36.4 286 1 Y137\_METJA  
41 40 36.4 292 1 YE95\_METJA  
42 40 36.4 326 1 YA55\_METJA  
43 40 36.4 472 1 PHR\_ECOLI  
44 40 36.4 572 1 PTLE\_STRAU  
45 40 36.4 578 1 VAC8\_YEAST

#### ALIGNMENTS

RESULT 1  
GTFC\_STRMU  
ID GTFC\_STRMU STANDARD; PRT: 1375 AA.  
AC P13470; P05427;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DE 15-DEC-1998 (Rel. 37, Last annotation update)  
DE GLUCOSYLTRANSFERASE-SI PRECURSOR (EC 2.4.1.5) (GTF-SI)  
DE (DEXTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE).  
GN GTF.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GS-5;  
RX MEDLINE=89137980; PubMed=2976010;  
RA Ueda S., Shiroza T., Kuramitsu H.K.;  
RT "Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.";  
RL Gene 69:101-109(1988).  
RN [2]  
RP SEQUENCE OF 1-349 FROM N.A.  
RC STRAIN=GS-5;  
RX MEDLINE=87308013; PubMed=3040685;  
RA Shiroza T., Ueda S., Kuramitsu H.K.;  
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";  
RL J. Bacteriol. 169:4263-4270(1987).  
CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE  
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -!- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) =  
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).  
CC -!- SUBCELLULAR LOCATION: SECRETED.  
CC -!- DISEASE: DENTAL CARIES.  
CC -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
CC FORMS OF GLUCANS.  
CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
CC BINDING PROTEIN FROM S.MUTANS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M22054; AAA8592.1; -;  
DR EMBL; M17361; AAA8589.1; -;  
DR PIR; J0345; J0345.  
DR PIR; C33135; C33135.  
DR InterPro; IPR002479; CW\_binding.



Query Match 95.5%; Score 105; DB 1; Length 1476;  
Best Local Similarity 95.5%; Pred. No. 1.1e-08;  
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPSYSFIRAHDSVQDLIRNII 22  
Best Local Similarity 95.5%; Score 105; DB 1; Length 1476;  
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 552 VPSYSFIRAHDSVQDLIRNII 573

RESULT 3

GTFL\_STRDO STANDARD; PRT; 1597 AA.

AC P11001;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-JUL-1989 (Rel. 11, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)

DE (SUCROSE 6-GLUCOSYLTRANSFERASE).

GTFI.

Streptococcus downei (Streptococcus sobrinus).

Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1317;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MFE28;

RA MEDLINE=87308014; PubMed=3040686;

RA Ferretti J.J., Gilpin M.L., Russell R.R.B.;

RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus sobrinus MFE28."

RL J. Bacteriol. 169:4271-4278(1987).

CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.

CC -!- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) = D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).

CC -!- SUBCELLULAR LOCATION: SECRETED.

CC -!- DISEASE: DENTAL CARIES.

CC -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF GLUCANS.

CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-BINDING PROTEIN FROM S.MUTANS.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL: M17391; AAC63063.1; .

DR InterPro: IPR002479; CW\_binding.

DR InterPro: IPR003318; Glyco\_hydro\_70.

DR Pfam: PF01473; CW\_binding\_1; 19.

DR Pfam: PF02324; Glyco\_hydro\_70; 1.

KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.

FT SIGNAL 1 38 POTENTIAL.

FT CHAIN 39 1597 GLUCOSYLTRANSFERASE-I.

FT DOMAIN 39 1050 CATALYTIC (APPROXIMATE).

FT DOMAIN 1099 1597 GLUCAN-BINDING (APPROXIMATE).

FT REPEAT 1099 1132 1.25 A, 2 B, AND 5 AC REPEATS.

FT REPEAT 1163 1213 A REPEAT.

FT REPEAT 1227 1277 AC REPEAT.

FT REPEAT 1292 1342 AC REPEAT.

FT REPEAT 1352 1399 B REPEAT.

FT REPEAT 1406 1455 AC REPEAT.

FT REPEAT 1465 1512 B REPEAT.

FT REPEAT 1519 1568 AC REPEAT.

FT REPEAT 1582 1597 A REPEAT (INCOMPLETE).

SQ SEQUENCE 1597 AA; 177080 MW; B9E86A200868798E CRC64;

Query Match 90.9%; Score 100; DB 1; Length 1597;  
Best Local Similarity 90.9%; Pred. No. 7.4e-08;  
Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSFIRAHDSVQDLIRNII 22  
Best Local Similarity 90.9%; Score 100; DB 1; Length 1597;  
Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 554 VPSYSFIRAHDSVQDLIRNII 575

RESULT 4

GTFL\_STRDO STANDARD; PRT; 1592 AA.

AC P27470;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)

DE (SUCROSE 6-GLUCOSYLTRANSFERASE).

OS Streptococcus downei (Streptococcus sobrinus).

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OX NCBI\_TaxID=1317;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=6715;

RC MEDLINE=91123227; PubMed=1704006;

RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K., Kagawa H.;

RA "Peptide sequences for sucrose splitting and glucan binding within Streptococcus sobrinus glucosyltransferase (water-insoluble glucan synthetase)."

RT J. Bacteriol. 173:989-996(1991).

CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.

CC -!- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) = D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).

CC -!- SUBCELLULAR LOCATION: SECRETED.

CC -!- DISEASE: DENTAL CARIES.

CC -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF GLUCANS.

CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-BINDING PROTEIN FROM S.MUTANS.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL: D90213; BAA14241.1; .

DR PIR: A38175; A38175.

DR HSP: P00695; 2HEE.

DR InterPro: IPR002479; CW\_binding.

DR InterPro: IPR003318; Glyco\_hydro\_70.

DR Pfam: PF01473; CW\_binding\_1; 16.

DR Pfam: PF02324; Glyco\_hydro\_70; 1.

KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.

FT SIGNAL 1 38 POTENTIAL.

FT CHAIN 39 1592 GLUCOSYLTRANSFERASE-I.

FT DOMAIN 39 1044 CATALYTIC (APPROXIMATE).

FT DOMAIN 1093 1592 GLUCAN-BINDING (APPROXIMATE).

FT DOMAIN 1093 1592 6.5 X TANDEM REPEATS.  
 FT REPEAT 1093 1142 1.  
 FT REPEAT 1158 1207 2.  
 FT REPEAT 1222 1272 3.  
 FT REPEAT 1287 1337 4.  
 FT REPEAT 1402 1451 5.  
 FT REPEAT 1514 1563 6.  
 FT REPEAT 1577 1592 7 (INCOMPLETE).  
 SQ SEQUENCE 1592 AA; 176167 MW; BC0A66D079351ECF CRC64;

Query Match 89.1%; Score 98; DB 1; Length 1592;  
 Best Local Similarity 86.4%; Pred. No. 1.5e-07;  
 Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 VPSYSFIRAHSEVDLIRNII 22  
 DB 548 VPSYSFIRAHSEVDIIRNII 569

RESULT 5  
 ID GTFS\_STRDO STANDARD; PRT; 1365 AA.  
 AC P29336;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 01-APR-1993 (Rel. 25, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)  
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
 GN GTF-S.  
 OS Streptococcus downei (Streptococcus sobrinus).  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1317;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MFE28;  
 RX MEDLINE=90316665; PubMed=2142479;  
 RA Gilmore K.S., Russell R.R., Ferretti J.J.;  
 RT "Analysis of the Streptococcus downei gtfS gene, which specifies a  
 glucosyltransferase that synthesizes soluble glucans.";  
 RL Infect. Immun. 58:2452-2458(1990).  
 CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
 AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -!- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =  
 D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -!- ENZYME REGULATION: GLUCAN SYNTHESIS BY GTF-S IS INDEPENDENT OF  
 PRIMER GLUCAN UNLIKE GTF-I.  
 CC -!- DISEASE: DENTAL CARIES.  
 CC -!- MISCELLANEOUS: GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA  
 1,6-GLUCOSE).  
 CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
 BINDING PROTEIN FROM S. MUTANS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M30943; AAA26898.1;  
 CC PIR; A41483; A41483.  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding.1; 10.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase; Glucosyltransferase; Signal; Repeat; Dental caries.  
 FT SIGNAL 1 36  
 FT CHAIN 37 1365 OR 37 (POTENTIAL).  
 FT CHAIN 37 1365 GLUCOSYLTRANSFERASE-S.

FT DOMAIN 37 1050 CATALYTIC (APPROXIMATE).  
 FT REPEAT 1083 1365 GLUCAN-BINDING (APPROXIMATE).  
 FT DOMAIN 1083 1365 4.5 X TANDEM REPEATS.  
 FT REPEAT 1083 1131 1.  
 FT REPEAT 1150 1199 2.  
 FT REPEAT 1225 1274 3.  
 FT REPEAT 1289 1339 4.  
 FT REPEAT 1353 1365 5 (INCOMPLETE).  
 SQ SEQUENCE 1365 AA; 151590 MW; 167296B5A2E8C476 CRC64;

Query Match 69.1%; Score 76; DB 1; Length 1365;  
 Best Local Similarity 72.7%; Pred. No. 0.00043;  
 Matches 16; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 VPSYSFIRAHSEVDLIRNII 22  
 DB 537 VPNVYFIRAHSEVDTRIAKII 558

RESULT 6  
 ID GTFD\_STRMU STANDARD; PRT; 1462 AA.  
 AC P49331; O69386; O69389; O69392; O69398;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)  
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
 GN GTFD.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GS-5;  
 RX MEDLINE=91100958; PubMed=2148600;  
 RA Honda O., Kato C., Kuramitsu H.K.;  
 RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding  
 the glucosyltransferase-S enzyme.";  
 RL J. Gen. Microbiol. 136:2099-2105(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MT4239, MT4245, MT4251, MT4467, AND MT8148;  
 RX MEDLINE=98231643; PubMed=9570124;  
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
 Kimura S., Hamada S.;  
 RT "Molecular analyses of glucosyltransferase genes among strains of  
 Streptococcus mutans.";  
 RL FEMS Microbiol. Lett. 161:331-336(1998).  
 CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
 AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -!- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =  
 D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -!- SUBCELLULAR LOCATION: SECRETED.  
 CC -!- DISEASE: DENTAL CARIES.  
 CC -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
 WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
 FORMS OF GLUCANS.  
 CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
 BINDING PROTEIN FROM S. MUTANS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M30943; AAA26898.1;  
 CC PIR; A41483; A41483.  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding.1; 10.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase; Glucosyltransferase; Signal; Repeat; Dental caries.  
 FT SIGNAL 1 36  
 FT CHAIN 37 1365 OR 37 (POTENTIAL).  
 FT CHAIN 37 1365 GLUCOSYLTRANSFERASE-S.



FT TRANSMEM 481 503 POTENTIAL.  
FT DOMAIN 504 540 CYTOPLASMIC (POTENTIAL).  
FT DISULFID 108 136 BY SIMILARITY.  
FT DISULFID 213 261 BY SIMILARITY.  
FT DISULFID 473 473 INTERCHAIN (BY SIMILARITY).  
FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 476 476 C -> SG (IN REF. 2).  
SQ SEQUENCE 540 AA; 59766 MW; 53EE0020A296D6F5 CRC64;

Query Match 43.6%; Score 48; DB 1; Length 540;  
Best Local Similarity 44.4%; Pred. No. 4.5;  
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 4 YSFIRAHSEVDLIRNI 21  
||| |:::|  
234 YSFDETKDLDELIVRNI 251

RESULT 8  
MAD2\_YEAST  
ID MAD2\_YEAST STANDARD; PRT; 196 AA.  
AC P40958;  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE MITOTIC SPINDLE CHECKPOINT COMPONENT MAD2 (MITOTIC MAD2 PROTEIN).  
GN MAD2 OR YJL030W OR J1256.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Li R., Havel C., Watson J.A., Murray A.W.;  
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Pohl T.M., Aljinovic G.;  
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP CHARACTERIZATION.  
RX MEDLINE=91330300; PubMed=1651172;  
RA Li R., Murray A.W.;  
RP "Feedback control of mitosis in budding yeast.";  
Cell 66:519-531(1991).  
RN [4]  
RP INTERACTIONS.  
RX MEDLINE=98128031; PubMed=9461437;  
RA Hwang L.H., Lau L.F., Smith D.L., Mistrot C.A., Hardwick K.G.,  
RA Hwang E.S., Amon A., Murray A.W.;  
RT "Budding yeast Cdc20: a target of the spindle checkpoint.";  
RL Science 279:1041-1044(1998).  
CC -1- FUNCTION: FEEDBACK CONTROL THAT PREVENTS CELLS WITH INCOMPLETELY  
ASSEMBLED SPINDLES FROM LEAVING MITOSIS.  
CC -1- SUBUNIT: THE SPINDLE CHECKPOINT COMPLEX IS COMPOSED OF MAD1, MAD2  
AND MAD3. IT INTERACTS WITH CDC20.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- SIMILARITY: BELONGS TO THE MAD2 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; 014132; AAA21385.1; -  
CC EMBL; Z49305; CAA89321.1; -  
CC PIR; S48302; S48302.

DR SGD: S0003567; MAD2.  
DR InterPro: IPR003511; HORMA.  
DR Pfam: PF02301; HORMA; 1.  
KW Cell cycle; Mitosis; Nuclear protein.  
SQ SEQUENCE 196 AA; 22284 MW; EFE59916C5720644 CRC64;

Query Match 41.8%; Score 46; DB 1; Length 196;  
Best Local Similarity 38.9%; Pred. No. 3.1;  
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

OY 5 SFIHSEVDLIRNI 22  
: : | | : | | |  
46 TLLKTHDDELKDYIRKIL 63

RESULT 9  
DUR1\_YEAST  
ID DUR1\_YEAST STANDARD; PRT; 1835 AA.  
AC P32528;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE UREA AMIDOLYASE [INCLUDES: UREA CARBOXYLASE (EC 6.3.4.6); ALLOPHANATE  
DE HYDROLASE (EC 3.5.1.54)].  
GN DUR1, 2 OR YBR208C OR YBR1448.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92199240; PubMed=1802034;  
RA Genbauffe F.S., Cooper T.G.;  
RT "The urea amidolyase (DUR1,2) gene of Saccharomyces cerevisiae.";  
RL DNA Seq. 2:19-32(1991).  
RN [2]  
RP SEQUENCE OF 1-893 FROM N.A.  
RC STRAIN=S288C;  
RA Rieger M.;  
RN Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 873-1835 FROM N.A.  
RC STRAIN=S288C;  
RA Feldmann H., Mannhaupt G., Schwarzlose C., Vetter I.;  
RN Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 1487-1835 FROM N.A.  
RC STRAIN=S288C;  
RX MEDLINE=93377417; PubMed=8368014;  
RA Bussereau F., Mallet L., Gaillon L., Jacquet M.;  
RT "A 12.8 kb segment, on the right arm of chromosome II from  
Saccharomyces cerevisiae including part of the DUR1,2 gene, contains  
five putative new genes.";  
RL Yeast 9:797-806(1993).  
CC -1- FUNCTION: HYDROLYSIS OF UREA TO AMMONIA AND CO2.  
CC -1- CATALYTIC ACTIVITY: ATP + UREA + CO(2) -> ADP + ORTHOPHOSPHATE +  
UREA-1-CARBOXYLATE.  
CC -1- CATALYTIC ACTIVITY: UREA-1-CARBOXYLATE + H(2)O -> 2 CO(2) +  
2 NH(3).  
CC -1- COFACTOR: BIOTIN.  
CC -1- PATHWAY: ALLANTOIN AND ARGININE METABOLISM.  
CC -1- SUBUNIT: MONOMER.  
CC -1- INDUCTION: BY ALLOPHANATE OR ITS NON-METABOLIZED ANALOG  
OXALURATE. REPRESSION IN THE PRESENCE OF READILY USED NITROGEN  
SOURCES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; 014132; AAA21385.1; -  
CC EMBL; Z49305; CAA89321.1; -  
CC PIR; S48302; S48302.

```

CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M64926; AAC1643.1; -
CC EMBL; Z36077; CAA85172.1; -
CC EMBL; Z21487; CAA79695.1; -
CC PIR; S46082; S46082.
CC HSSP; P24182; IBNC.
CC SGD; S0000412; DURL1.2.
CC InterPro; IPR000120; Amidase.
CC InterPro; IPR001882; Biotin.
CC InterPro; IPR000089; Biotin_lipoyl.
CC InterPro; IPR000901; CPSase.
CC InterPro; IPR003778; DUF183.
CC Pfam; PF01425; Amidase; 1.
CC Pfam; PF00364; biotin_lipoyl; 1.
CC Pfam; PF00289; CPSase_L_chain; 1.
CC Pfam; PF02626; DUF183; 1.
CC PROSITE; PS00188; BIOTIN; 1.
CC PROSITE; PS00866; CPSASE_1; 1.
CC PROSITE; PS00867; CPSASE_2; 1.
CC Biotin; Hydrolase; Multifunctional enzyme; Arginine metabolism;
CC Biotin; ATP-binding.
CC NP_BIND 122 129 ATP (POTENTIAL).
CC BINDING 1798 1798 BIOTIN.
CC SIMILAR 1779 1835 TO OTHER BIOTIN CARBOXYLASES C-TERMINI.
CC CONFLICT 96 96 P -> R (IN REF. 1).
CC CONFLICT 256 258 LKK -> KKN (IN REF. 1).
CC CONFLICT 459 459 I -> M (IN REF. 1).
CC CONFLICT 830 830 E -> K (IN REF. 1).
CC CONFLICT 1395 1395 D -> E (IN REF. 1).
CC SEQUENCE 1835 AA; 201831 MW; F52B0DD0FE42CD65 CRC64;

Query Match 40.9%; Score 45; DB 1; Length 1835;
Best Local Similarity 42.1%; Pred. No. 53;
Matches 8; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 2 PSYFFIRADSEVQDLIRN 20
|::: |::|::|
Db 102 PSFAYEPSKDSKVVELLRN 120

RESULT 10
ID BACC_BACLI STANDARD; PRT; 6359 AA.
AC O68008;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE BACITRACIN SYNTHETASE 3 (BA3) [INCLUDES: ATP-DEPENDENT ISOLEUCINE
DE ADENYLASE (ILEA) (ISOLEUCINE ACTIVASE); ATP-DEPENDENT D-PHENYLALANINE
DE ADENYLASE (D-PHEA) (D-PHENYLALANINE ACTIVASE); ATP-DEPENDENT HISTIDINE
DE ADENYLASE (HISA) (HISTIDINE ACTIVASE); ATP-DEPENDENT D-ASPARTATE
DE ADENYLASE (D-ASPA) (D-ASPARTATE ACTIVASE); ATP-DEPENDENT ASPARAGINE
DE ADENYLASE (ASNA) (ASPARAGINE ACTIVASE); ASPARTATE RACEMASE
DE (EC 5.1.1.13); PHENYLALANINE RACEMASE [ATP HYDROLYSING]
DE (EC 5.1.1.11)].
GN BACC.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10716;
RX MEDLINE=98089193; PubMed=9427658;
RA Konz D., Klenz A., Schoergerdorfer K., Marahiel M.A.;
RT "The bacitracin biosynthesis operon of Bacillus licheniformis ATCC
RT 10716: molecular characterization of three multi-modular peptide
RT synthetases.";
RL Chem. Biol. 4:927-937(1997).
CC -1- FUNCTION: INDUCES PEPTIDE SYNTHESIS, ACTIVATES AND INCORPORATES
CC FIVE AMINO ACIDS, FORMS A THIAZOLINE RING BETWEEN THE FIRST TWO

```

```

CC AMINO ACIDS AND INCORPORATES A D-GLUTAMINE IN THE FOURTH POSITION.
CC -1- CATALYTIC ACTIVITY: L-ASPARTATE = D-ASPARTATE.
CC + D-PHENYLALANINE.
CC -1- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC (POTENTIAL).
CC -1- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF THE CYCLIC PEPTIDE
CC ANTIBIOTIC BACITRACIN.
CC -1- SUBUNIT: LARGE MULTISUBUNIT COMPLEX OF BAI, BAI2 AND BAI3.
CC -1- DOMAIN: CONSISTS OF FIVE MODULES WITH TWO EPIMERIZATION DOMAINS IN
CC THE SECOND AND FOURTH MODULES, AND A PUTATIVE C-TERMINAL
CC THIOESTERASE DOMAIN. EACH MODULE INCORPORATES ONE AMINO ACID INTO
CC THE PEPTIDE PRODUCT AND CAN BE FURTHER SUBDIVIDED INTO DOMAINS
CC RESPONSIBLE FOR SUBSTRATE ADENYLATION, THIOLIZATION, CONDENSATION
CC (NOT FOR THE INITIATION MODULE), AND EPIMERIZATION (OPTIONAL), AND
CC N METHYLATION (OPTIONAL).
CC -1- MISCELLANEOUS: BACITRACIN IS A MIXTURE OF AT LEAST TEN CYCLIC
CC DODECAPEPTIDES, THAT DIFFER BY ONE OR TWO AMINO ACIDS. THE MOST
CC ABUNDANT IS BACITRACIN A, A BRANCHED CYCLIC DODECAPEPTIDE. IT
CC CONTAINS AN N-TERMINAL LINEAR PENTAPEPTIDE MOIETY (ILE-CYS-LEU-D-
CC GLU-ILE) WITH AN ISOLEUCINE-CYSTEINE THIAZOLINE CONDENSATION
CC PRODUCT AND A C-TERMINAL HEPTAPEPTIDE RING (LYS-D-ORN-ILE-D-PHE-
CC HIS-D-ASP-ASN), IN WHICH THE FREE ALPHA-CARBOXY GROUP OF THE C-
CC TERMINAL ASN IS BOUND TO THE EPSILON-AMINO GROUP OF LYSINE. IT
CC CONTAINS FOUR AMINO ACIDS IN THE D-CONFIGURATION (GLU-4, ORN-7,
CC PHE-9, AND ASP-11).
CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF007865; AAC06348.1; -.
CC InterPro; IPR000873; AMP-bind.
CC InterPro; IPR001242; DUF4.
CC InterPro; IPR003880; Phosphopant_attach.
CC InterPro; IPR001031; Thioesterase.
CC Pfam; PF00501; AMP-binding; 5.
CC Pfam; PF00668; Condensation; 7.
CC Pfam; PF00550; pp-binding; 5.
CC Pfam; PF00975; Thioesterase; 1.
CC PRINTS; PR00154; AMPBINDING.
CC PROSITE; PS00012; PHOSPHOPANTETHEINE; 4.
CC PROSITE; PS00455; AMP_BINDING; 5.
CC PROSITE; PS00075; ACP_DOMAIN; 5.
CC Ligase; Isomerase; Hydrolase; Antibiotic biosynthesis;
CC Phosphopantetheine; Multifunctional enzyme; Repeat.
CC REPEAT 461 1034 DOMAIN 1 (ISOLEUCINE-ACTIVATING).
CC REPEAT 1517 2064 DOMAIN 2 (D-PHENYLALANINE-ACTIVATING).
CC REPEAT 2999 3570 DOMAIN 3 (HISTIDINE-ACTIVATING).
CC REPEAT 4047 4612 DOMAIN 4 (D-ASPARTIC ACID-ACTIVATING).
CC REPEAT 5549 6129 DOMAIN 5 (ASPARAGINE-ACTIVATING).
CC REPEAT 966 1034 ACYL CARRIER (ACP) 1.
CC DOMAIN 1998 2064 ACYL CARRIER (ACP) 2.
CC DOMAIN 3502 3570 ACYL CARRIER (ACP) 3.
CC DOMAIN 4544 4612 ACYL CARRIER (ACP) 4.
CC DOMAIN 6052 6129 ACYL CARRIER (ACP) 5.
CC BINDING 996 996 PHOSPHOPANTETHEINE (BY SIMILARITY).
CC BINDING 2028 2028 PHOSPHOPANTETHEINE (BY SIMILARITY).
CC BINDING 3532 3532 PHOSPHOPANTETHEINE (BY SIMILARITY).
CC BINDING 4574 4574 PHOSPHOPANTETHEINE (BY SIMILARITY).
CC BINDING 6082 6082 PHOSPHOPANTETHEINE (BY SIMILARITY).
CC SEQUENCE 6359 AA; 722923 MW; 82A273C546253074 CRC64;

Query Match 40.9%; Score 45; DB 1; Length 6359;
Best Local Similarity 45.5%; Pred. No. 2.1e+02;
Matches 10; Conservative 4; Mismatches 4; Indels 4; Gaps 1;

```

Bacillus subtilis.  
Bacteria; Firmicutes; Bacillus/Clostridium group;  
Bacillus/Staphylococcus group; Bacillus.



RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / FY1679;  
 RX MEDLINE=96367601; PubMed=8771715;  
 RA Garcia-Cantalejo J.M., Boskovic J., Jimenez A.;  
 RT "Sequence analysis of a 14.2 kb fragment of *Saccharomyces cerevisiae*  
 chromosome XIV that includes the *Ypf53*, *TRNA*Leu and *gsr m2* genes and  
 RT four new open reading frames.";  
 RL Yeast 12:599-608(1996).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X85811; CAA59823.1; ALT\_INIT.  
 DR EMBL; 271370; CAA95970.1;  
 DR SGD; S0005038; YNL094W.  
 DR Hypothetical protein.  
 DR SEQUENCE 587 AA; 66134 MW; B85C525548BA34BC CRC64;  
 CC -----  
 CC Query Match 40.0%; Score 44; DB 1; Length 587;  
 CC Best Local Similarity 45.0%; Pred. No. 22;  
 CC Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
 CC -----  
 QY 3 SYSFIRAHDSVQDLIRNII 22  
 Db 430 SMSDVPSHDEVMEVNNII 449  
 | | | : || | | : || |  
 RESULT 14  
 P11A\_BOVIN STANDARD; PRT; 1068 AA.  
 ID P11A\_BOVIN STANDARD; PRT; 1068 AA.  
 AC P32871;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM  
 DE (EC 2.7.1.137) (PI3-KINASE P110 SUBUNIT ALPHA) (PTDINS-3-KINASE P110)  
 DE (PI3K).  
 GN PI3KA.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OC NCBI\_TaxID=9913;  
 OX [1]  
 SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 TISSUE=Brain;  
 RX MEDLINE=92334059; PubMed=1322797;  
 RA Hiles I.D., Otsu M., Volinia S., Fry M.J., Gout I., Dhand R.,  
 RA Panayotou G., Ruiz-Larrea F., Thompson A., Totty N.F., Hsuan J.J.,  
 RA Courtneidge S.A., Parker P.J., Waterfield M.D.;  
 RT "Phosphatidylinositol 3-kinase: structure and expression of the 110  
 RT kd catalytic subunit".  
 RL Cell 70:419-429(1992).  
 CC -1- FUNCTION: PHOSPHORYLATES PTDINS, PTDINS4P AND PTDINS(4,5)P2 WITH A  
 CC PREFERENCE FOR PTDINS(4,5)P2.  
 CC -1- CATALYTIC ACTIVITY: ATP + 1-PHOSPHATIDYL-ID-MYO-INOSITOL = ADP +  
 CC 1-PHOSPHATIDYL-ID-MYO-INOSITOL 3-PHOSPHATE.  
 CC -1- SUBUNIT: HETERODIMER OF A P110 (CATALYTIC) AND A P85 (REGULATORY)  
 CC SUBUNIT.  
 CC -1- SIMILARITY: BELONGS TO THE PI3/PI4-KINASES FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 C2 DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M93252; AAA30698.1;  
 DR PIR; A43322; A43322.  
 DR InterPro; IPR000008; C2.  
 DR InterPro; IPR002420; PI3K\_C2.  
 DR InterPro; IPR003113; PI3K\_p85B.  
 DR InterPro; IPR000341; PI3K\_ras\_bind.  
 DR InterPro; IPR001263; PI3Ka.  
 DR InterPro; IPR000403; PI3\_PI4\_kinase.  
 DR Pfam; PF00613; PI3Ka; 1.  
 DR Pfam; PF00792; PI3K\_C2; 1.  
 DR Pfam; PF02192; PI3K\_p85B; 1.  
 DR Pfam; PF00794; PI3K\_rbd; 1.  
 DR Pfam; PF00454; PI3\_PI4\_kinase; 1.  
 DR SMART; SM00239; C2; 1.  
 DR SMART; SM00145; PI3Ka; 1.  
 DR SMART; SM00146; PI3KC; 1.  
 DR SMART; SM00142; PI3K\_C2; 1.  
 DR SMART; SM00143; PI3K\_p85B; 1.  
 DR SMART; SM00144; PI3K\_rbd; 1.  
 DR PROSITE; PS00004; C2\_DOMAIN\_2; FALSE\_NEG.  
 DR PROSITE; PS00915; PI3\_4\_KINASE\_1; 1.  
 DR PROSITE; PS00916; PI3\_4\_KINASE\_2; 1.  
 DR PROSITE; PS02090; PI3\_4\_KINASE\_3; 1.  
 KW Transferase; Kinase; Multigene family.  
 FT DOMAIN 319 428 C2 DOMAIN.  
 FT DOMAIN 797 1068 PI3K/PI4K.  
 SQ SEQUENCE 1068 AA; 124327 MW; C753DCC2F39FDDF0 CRC64;  
 CC -----  
 CC Query Match 40.0%; Score 44; DB 1; Length 1068;  
 CC Best Local Similarity 45.5%; Pred. No. 42;  
 CC Matches 10; Conservative 2; Mismatches 10; Indels 0; Gaps 0;  
 CC -----  
 QY 1 VPSYSFIRAHDSVQDLIRNII 22  
 Db 123 MPVCFEDVMKDPQDFRRNII 144  
 | | | : || | | : || |  
 RESULT 15  
 P11A\_HUMAN STANDARD; PRT; 1068 AA.  
 ID P11A\_HUMAN STANDARD; PRT; 1068 AA.  
 AC P42336; Q9762;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM  
 DE (EC 2.7.1.137) (PI3-KINASE P110 SUBUNIT ALPHA) (PTDINS-3-KINASE P110)  
 DE (PI3K).  
 GN PI3KA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=95229146; PubMed=7713498;  
 RA Volinia S., Hiles I., Ormordroyd E., Nizetic D., Antonacci R.,  
 RA Rocchi M., Waterfield M.;  
 RT "Molecular cloning, cDNA sequence, and chromosomal localization of  
 RT the human phosphatidylinositol 3-kinase p110 alpha (PIK3CA) gene".  
 RL Genomics 24:472-477(1994).  
 CC [2]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=97196568; PubMed=9043658;  
 RA Stirling S.M., Ahern J., Conroy R.R., Barnett S.F., Ledder L.M.,  
 RA Oliff A., Heimbrock D.C.;  
 RT "Cloning and mutagenesis of the p110 alpha subunit of human  
 RT phosphoinositide 3'-hydroxylase".  
 RL Bioorg. Med. Chem. 5:65-74(1997).  
 CC -1- FUNCTION: PHOSPHORYLATES PTDINS, PTDINS4P AND PTDINS(4,5)P2 WITH A

```
CC      PREFERENCE FOR PTAINS(4,5)P2.
CC      -|- CATALYTIC ACTIVITY: ATP + 1-PHOSPHATIDYL-1D-MYO-INOSITOL = ADP +
CC      1-PHOSPHATIDYL-1D-MYO-INOSITOL 3-PHOSPHATE.
CC      -|- SUBUNIT: HETERODIMER OF A P110 (CATALYTIC) AND A P85 (REGULATORY)
CC      SUBUNIT.
CC      -|- SIMILARITY: BELONGS TO THE PI3/PI4-KINASES FAMILY.
CC      -|- SIMILARITY: CONTAINS 1 C2 DOMAIN.
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; Z29090; CAA82333.1; -
CC      EMBL; U79143; AAB39753.1; -
CC      MIM; 171834; -
CC      InterPro; IPR000008; C2.
CC      InterPro; IPR002420; PI3K_C2
CC      InterPro; IPR003113; PI3K_P85B.
CC      InterPro; IPR000341; PI3K_ras_bind.
CC      InterPro; IPR001263; PI3Ka.
CC      InterPro; IPR000403; PI3_P14_kinase.
CC      Pfam; PF006113; PI3Ka; 1.
CC      Pfam; PF00792; PI3K_C2; 1.
CC      Pfam; PF02192; PI3K_P85B; 1.
CC      Pfam; PF00794; PI3K_rbd; 1.
CC      Pfam; PF00454; PI3_P14_kinase; 1.
CC      SMART; SM00239; C2; 1.
CC      SMART; SM00145; PI3Ka; 1.
CC      SMART; SM00146; PI3K; 1.
CC      SMART; SM00142; PI3K_C2; 1.
CC      SMART; SM00143; PI3K_P85B; 1.
CC      SMART; SM00144; PI3K_rbd; 1.
CC      PROSITE; PS50004; C2_DOMAIN_2; FALSE_NEG.
CC      PROSITE; PS00915; PI3_4_KINASE_1; 1.
CC      PROSITE; PS00916; PI3_4_KINASE_2; 1.
CC      PROSITE; PS0290; PI3_4_KINASE_3; 1.
CC      TRANSFERASE; Kinase; Multigene family.
CC      DOMAIN 319 428 C2 DOMAIN.
CC      FT DOMAIN 797 1068 PI3K/PI4K.
CC      FT CONFLICT 43 43 V -> I (IN REF. 2).
CC      FT CONFLICT 170 170 H -> N (IN REF. 2).
CC      FT CONFLICT 187 187 R -> K (IN REF. 2).
CC      FT CONFLICT 286 287 KM -> ML (IN REF. 2).
CC      FT CONFLICT 332 332 R -> S (IN REF. 2).
CC      FT CONFLICT 346 346 L -> V (IN REF. 2).
CC      FT CONFLICT 723 723 R -> K (IN REF. 2).
CC      FT CONFLICT 751 751 L -> F (IN REF. 2).
CC      FT CONFLICT 767 767 K -> E (IN REF. 2).
CC      SQ SEQUENCE 1068 AA; 124412 MW; 9E16BA7401A87B57 CRC64;
```

```
Query Match 40.0%; Score 44; DB 1; Length 1068;
Best Local Similarity 45.5%; Pred. No. 42;
Matches 10; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

Qy 1 VPSYSFIRAHDSVQDLIRNII 22
   :| | | | | | | |
Db 123 MPVCFDMVKDPEYQDFRRNII 144
```

Search completed: March 27, 2002, 14:27:06  
Job time: 1648 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:26:07 ; Search time 188.53 Seconds  
(without alignments)  
17.069 Million cell updates/sec

Title: US-09-290-049a-15

Perfect score: 110  
Sequence: 1 VPSYSFIRAHSEVQDLIRNII 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	110	100.0	1455	2	069388
2	105	95.5	1390	2	069385
3	105	95.5	1455	2	069382
4	105	95.5	1455	2	069391
5	105	95.5	1455	2	069397
6	98	89.1	1590	2	Q5263
7	98	89.1	1590	2	Q5983
8	83	75.5	1290	2	Q48756
9	83	75.5	1477	2	Q9L466
10	83	75.5	1508	2	Q52224
11	83	75.5	1508	2	Q9EZH5
12	82	74.5	1527	2	Q9ZAR4
13	80	72.7	1016	2	Q9LCJ7
14	77	70.0	1512	2	Q9WJ35
15	76	69.1	1338	2	Q9WXJ4
16	71	64.5	1449	2	Q68542
17	71	64.5	1449	2	Q55264
18	71	64.5	2057	2	Q9RE05
19	70	63.6	1577	2	Q55265

20	68	61.8	1518	2	Q00600
21	68	61.8	1575	2	Q9LCH3
22	68	61.8	1577	2	Q54178
23	68	61.8	1599	2	Q00599
24	50	45.5	51	2	Q9RNP7
25	50	45.5	93	2	Q9ZIX9
26	46.5	42.3	414	10	Q65491
27	46	41.8	66	2	Q9K5K6
28	46	41.8	597	2	Q9PR58
29	46	41.8	1032	10	Q9RTJ9
30	45.5	41.4	767	3	Q74759
31	45.5	41.4	956	2	Q9PJF0
32	45	40.9	136	2	Q31949
33	45	40.9	136	9	Q64072
34	45	40.9	852	13	Q9PTE5
35	44.5	40.5	954	2	Q84594
36	44	40.0	202	1	Q9YDI8
37	44	40.0	327	2	Q9X303
38	44	40.0	327	10	Q9C7I1
39	44	40.0	371	8	Q9B8A8
40	44	40.0	525	10	Q9ATU4
41	44	40.0	567	12	Q9DHS5
42	44	40.0	734	2	Q9KNO3
43	44	40.0	1068	13	Q42391
44	44	40.0	1083	12	Q39483
45	43.5	39.5	273	10	Q9M7X2

ALIGNMENTS

RESULT 1

O69388	PRELIMINARY;	PRT; 1455 AA.
ID O69388		
AC O69388		
DT 01-AUG-1998 (TrEMBLrel. 07, Created)		
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)		
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)		
DE GLUCOSYLTRANSFERASE-SI.		
GN GTF.		
OS Streptococcus mutans.		
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;		
OC Streptococcus.		
OX NCBI_TaxID=1209;		
RN [1]		
RP SEQUENCE FROM N.A.		
RC STRAIN=MT4239;		
RX MEDLINE=98231643; PubMed=9570124;		
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,		
RA Kimura S., Hamada S.;		
RT "Molecular analyses of glucosyltransferase genes among strains of		
RT Streptococcus mutans."		
RL FEMS Microbiol. Lett. 161:331-336(1998).		
DR EMBL; D88652; RAA26110.1;		
DR InterPro; IPR002479; CW_binding.		
DR InterPro; IPR003318; Glyco_hydro_70.		
DR Pfam; PF01473; CW_binding_1; 10.		
DR Pfam; PF02324; Glyco_hydro_70; 1.		
KW Transferase.		
SQ SEQUENCE 1455 AA; 163045 MW; 6D90A4978D35DD82 CRC64;		

Query Match 100.0%; Score 110; DB 2; Length 1455;  
Best Local Similarity 100.0%; Pred. No. 2.2e-08;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPSYSFIRAHSEVQDLIRNII 22

Db 578 VPSYSFIRAHSEVQDLIRNII 599

RESULT 2

O69385

ID O69385 PRELIMINARY; PRT; 1390 AA.  
AC O69385;  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTFC.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MT4245;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
RA Kimura S., Hamada S.;  
"Molecular analyses of glucosyltransferase genes among strains of  
Streptococcus mutans."  
FEMS Microbiol. Lett. 161:331-336(1998).  
DR EMBL; D88655; BAA26106.1; -.  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 7.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1390 AA; 155375 MW; 8847E4956EF05E9F CRC64;

Query Match 95.5%; Score 105; DB 2; Length 1390;  
Best Local Similarity 95.5%; Pred. No. 1.2e-07;  
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPSYSFIRAHDSVDLIRNII 22  
Db 578 VPSYSFIRAHDSVDLIRNII 599  
|||||

RESULT 3  
ID O69382 PRELIMINARY; PRT; 1455 AA.  
AC O69382;  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTFC.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MT8148;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
RA Kimura S., Hamada S.;  
"Molecular analyses of glucosyltransferase genes among strains of  
Streptococcus mutans."  
FEMS Microbiol. Lett. 161:331-336(1998).  
DR EMBL; D88652; BAA26102.1; -.  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 9.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1455 AA; 162969 MW; 27D4D3A1EECA2939 CRC64;

Query Match 95.5%; Score 105; DB 2; Length 1455;  
Best Local Similarity 95.5%; Pred. No. 1.3e-07;  
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPSYSFIRAHDSVDLIRNII 22  
Db 578 VPSYSFIRAHDSVDLIRNII 599  
|||||

RESULT 4  
ID O69391 PRELIMINARY; PRT; 1455 AA.  
AC O69391;  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTFC.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MT4251;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
RA Kimura S., Hamada S.;  
"Molecular analyses of glucosyltransferase genes among strains of  
Streptococcus mutans."  
FEMS Microbiol. Lett. 161:331-336(1998).  
DR EMBL; D88661; BAA26114.1; -.  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 9.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1455 AA; 162804 MW; 683A359D873E9E1A CRC64;

Query Match 95.5%; Score 105; DB 2; Length 1455;  
Best Local Similarity 95.5%; Pred. No. 1.3e-07;  
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPSYSFIRAHDSVDLIRNII 22  
Db 578 VPSYSFIRAHDSVDLIRNII 599  
|||||

RESULT 5  
ID O69397 PRELIMINARY; PRT; 1455 AA.  
AC O69397;  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTFC.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MT4467;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
RA Kimura S., Hamada S.;  
"Molecular analyses of glucosyltransferase genes among strains of  
Streptococcus mutans."  
FEMS Microbiol. Lett. 161:331-336(1998).  
DR EMBL; D89978; BAA26120.1; -.  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 9.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.

SO SEQUENCE 1455 AA; 162913 MW; A1263427BF24E8E1 CRC64;

Query Match 95.5%; Score 105; DB 2; Length 1455;

Best Local Similarity 95.5%; Pred. No. 1.3e-07;

Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPSYSFIRAHDSVDLIRNII 22

||||| ||||||| ||||||| |||||||

Db 578 VPSYSFIRAHDSVDLIRNII 599

RESULT 6

ID Q55263 PRELIMINARY; PRT; 1590 AA.

AC Q55263;

DT 01-NOV-1996 (TReMBLrel. 01, Created)

DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)

DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)

DE GLUCOSYLTRANSFERASE GTF-I.

Streptococcus sobrinus.

Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

Streptococcus.

NCBI\_TaxID=1310;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 33478;

RA Sato S.;

RL Ann. Kagoshima Univ. Dental School 16:23-29(1996).

DR EMBL; D63570; BAA09792.1; -

DR InterPro; IPR002479; CW\_binding.

DR InterPro; IPR003318; Glyco\_hydro\_70.

DR Pfam; PF01473; CW\_binding.1; 15.

DR Pfam; PF02324; Glyco\_hydro\_70; 1.

KW Transferase.

SO SEQUENCE 1590 AA; 176057 MW; 9DF7A3F2C6E4FD43 CRC64;

Query Match

Best Local Similarity 89.1%; Score 98; DB 2; Length 1590;

Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSFIRAHDSVDLIRNII 22

||||| ||||||| ||||||| |||||||

Db 548 VPSYSFIRAHDSVDLIRNII 569

RESULT 7

Q59983 PRELIMINARY; PRT; 1590 AA.

Q59983;

DT 01-NOV-1996 (TReMBLrel. 01, Created)

DT 01-JAN-1998 (TReMBLrel. 01, Last sequence update)

DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)

DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE

6-GLUCOSYLTRANSFERASE).

Streptococcus.

NCBI\_TaxID=1310;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=OMZ176;

RA MEDLINE=94146405; PubMed=8312602;

RT Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;

"DNA sequence of the glucosyltransferase gene of serotype d

Streptococcus sobrinus.";

DNA Seq. 4:19-27(1993).

CC -I- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) = D-

FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).

EMBL; D13858; BAA02976.1; -

DR InterPro; IPR002479; CW\_binding.

DR InterPro; IPR003318; Glyco\_hydro\_70.

DR Pfam; PF01473; CW\_binding.1; 16.

DR Pfam; PF02324; Glyco\_hydro\_70; 1.

KW Signal; Transferase; Glycosyltransferase.

FT SIGNAL 1 38 POTENTIAL.

FT CHAIN 39 1590 GLUCOSYLTRANSFERASE-I.

SO SEQUENCE 1590 AA; 175955 MW; C3C83A57CF3C2B0E CRC64;

Query Match

Best Local Similarity 89.1%; Score 98; DB 2; Length 1590;

Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSFIRAHDSVDLIRNII 22

||||| ||||||| ||||||| |||||||

Db 548 VPSYSFIRAHDSVDLIRNII 569

RESULT 8

ID Q48756 PRELIMINARY; PRT; 1290 AA.

AC Q48756;

DT 01-NOV-1996 (TReMBLrel. 01, Created)

DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)

DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)

DE DEXTRANSUCRASE.

OS Leuconostoc mesenteroides.

Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;

Leuconostoc.

NCBI\_TaxID=1245;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NRRL B1299;

RA MEDLINE=97136686; PubMed=8982063;

RT Monchois V., Willemot R.M., Remaud-Simeon M., Croux C., Monsan P.;

"Cloning and sequencing of a gene coding for a novel dextranucrase

from Leuconostoc mesenteroides NRRL B-1299 synthesizing only alpha (1-

6) and alpha (1-3) linkages.";

RL Gene 182:23-32(1996).

DR EMBL; U38181; AAB40875.1; -

DR InterPro; IPR002479; CW\_binding.

DR InterPro; IPR003318; Glyco\_hydro\_70.

DR Pfam; PF01473; CW\_binding.1; 11.

DR Pfam; PF02324; Glyco\_hydro\_70; 1.

SO SEQUENCE 1290 AA; 145590 MW; 3555C2E96B749FAA CRC64;

Query Match

Best Local Similarity 75.5%; Score 83; DB 2; Length 1290;

Matches 16; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 PSYGFIRAHDSVDLIRNII 22

||||| ||||||| ||||||| |||||||

Db 388 PNYSFIRAHDSVDLIRNII 408

RESULT 9

ID Q9L466 PRELIMINARY; PRT; 1477 AA.

AC Q9L466;

DT 01-OCT-2000 (TReMBLrel. 15, Created)

DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)

DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)

DE DEXTRANSUCRASE (EC 2.4.1.5).

GN DSRC.

OS Leuconostoc mesenteroides.

Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;

Leuconostoc.

NCBI\_TaxID=1245;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NRRL B-1355;

RA Arguello-Morales M.A., Remaud-Simeon M., Pizzut S., Sarcabal P.,

```
RA Willemot R.M., Monsan P.;
RT "Sequence analysis of the gene encoding alternansucrase, a sucrose
RL glucosyltransferase from Leuconostoc mesenteroides NRRL B-1355.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ250172; CAB76565.1; -.
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1; 14.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 1477 AA; 164886 MW; E6F5710DEDFCB831 CRC64;

Query Match 75.5%; Score 83; DB 2; Length 1477;
Best Local Similarity 68.2%; Pred. No. 0.00036;
Matches 15; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPSYSFIRAHDSVQDLIRNII 22
:||||:|||||||:| ||
603 IPNYSFVRAHDSVQTVIAQII 624

RESULT 10
O52224 PRELIMINARY; PRT; 1508 AA.
ID O52224
AC O52224;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE 6-
DE GLUCOSYLTRANSFERASE).
GN DSRB.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL B-1299;
RA Monchois V., Renaud-Simeon M., Monsan P., Willemot R.M.;
RL FEMS Microbiol. Lett. 0:0-0(1998).
CC -1- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) = D-
CC FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).
DR EMBL: AF030129; AAB95453.1; -.
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1; 14.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 1508 AA; 168511 MW; E70CECB57A70D1F0 CRC64;

Query Match 75.5%; Score 83; DB 2; Length 1508;
Best Local Similarity 68.2%; Pred. No. 0.00036;
Matches 15; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPSYSFIRAHDSVQDLIRNII 22
:||||:|||||||:| ||
634 IPNYSFVRAHDSVQTVIAQII 655

RESULT 11
O9EZH5 PRELIMINARY; PRT; 1508 AA.
ID O9EZH5
AC O9EZH5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DEXTRANSUCRASE DSRB742.
GN DSRB742.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Leuconostoc.

RA Willemot R.M., Monsan P.;
RT "Sequence analysis of the gene encoding alternansucrase, a sucrose
RL glucosyltransferase from Leuconostoc mesenteroides NRRL B-1355.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ250172; CAB76565.1; -.
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1; 14.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 1477 AA; 164886 MW; E6F5710DEDFCB831 CRC64;

Query Match 75.5%; Score 83; DB 2; Length 1477;
Best Local Similarity 68.2%; Pred. No. 0.00036;
Matches 15; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPSYSFIRAHDSVQDLIRNII 22
:||||:|||||||:| ||
603 IPNYSFVRAHDSVQTVIAQII 624

RESULT 10
O52224 PRELIMINARY; PRT; 1508 AA.
ID O52224
AC O52224;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE 6-
DE GLUCOSYLTRANSFERASE).
GN DSRB.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL B-1299;
RA Monchois V., Renaud-Simeon M., Monsan P., Willemot R.M.;
RL FEMS Microbiol. Lett. 0:0-0(1998).
CC -1- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) = D-
CC FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).
DR EMBL: AF030129; AAB95453.1; -.
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1; 14.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 1508 AA; 168511 MW; E70CECB57A70D1F0 CRC64;

Query Match 75.5%; Score 83; DB 2; Length 1508;
Best Local Similarity 68.2%; Pred. No. 0.00036;
Matches 15; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPSYSFIRAHDSVQDLIRNII 22
:||||:|||||||:| ||
634 IPNYSFVRAHDSVQTVIAQII 655

RESULT 11
O9EZH5 PRELIMINARY; PRT; 1508 AA.
ID O9EZH5
AC O9EZH5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DEXTRANSUCRASE DSRB742.
GN DSRB742.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Leuconostoc.
```

```
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B-742CB;
RA Kim H.-S., Kim D., Ryu H.-J., Robyt J.F.;
RT "Leuconostoc mesenteroides B-742CB, a dextranucrase gene.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF294469; AAG38021.1; -.
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1; 14.
DR Pfam: PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1508 AA; 168542 MW; E2FCFA0F87AE4F3A CRC64;

Query Match 75.5%; Score 83; DB 2; Length 1508;
Best Local Similarity 68.2%; Pred. No. 0.00036;
Matches 15; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPSYSFIRAHDSVQDLIRNII 22
:||||:|||||||:| ||
634 IPNYSFVRAHDSVQTVIAQII 655

RESULT 12
O9ZAR4 PRELIMINARY; PRT; 1527 AA.
ID O9ZAR4
AC O9ZAR4;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DEXTRANSUCRASE.
GN DEX.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL B-512-F;
RA Bhatnagar R., Singh D.K.S.;
RT "Cloning and Molecular Characterization of Dextranucrase Gene from
RT Leuconostoc mesenteroides NRRL B-512F.";
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U81374; AAD10952.1; -.
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1; 16.
DR Pfam: PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1527 AA; 169709 MW; 1DFAFA237C743398 CRC64;

Query Match 74.5%; Score 82; DB 2; Length 1527;
Best Local Similarity 63.6%; Pred. No. 0.00052;
Matches 14; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPSYSFIRAHDSVQDLIRNII 22
:||||:|||||||:| ||
652 IPNYSFVRAHDSVQTVIAQII 673

RESULT 13
O9LCJ7 PRELIMINARY; PRT; 1016 AA.
ID O9LCJ7
AC O9LCJ7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DEXTRANSUCRASE.
GN DSRT.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Leuconostoc.
```

OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL B-512F;  
 RX MEDLINE=20169623; PubMed=10705445;  
 RA Funane K., Mizuno K., Takahara H., Kobayashi M.;  
 RT "Gene encoding a dextranucrase-like protein in *Leuconostoc*  
 RL mesenteroides NRRL B-512F.";  
 RL Biosci. Biotechnol. Biochem. 64:29-38(2000).  
 DR EMBL: AB020020; BAA90527.1; -  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 SQ SEQUENCE 1016 AA; 110343 MW; 8896FEDE13CCCB47 CRC64;

Query Match 72.7%; Score 80; DB 2; Length 1016;  
 Best Local Similarity 71.4%; Pred. No. 0.00069;  
 Matches 15; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Oy 2 PSYSFIRAHDSVDLIRNII 22  
 I:|||||:|||||:|  
 625 PNYSEVRAHDSVQTVAIEII 645

RESULT 14  
 Q9WXJ5 PRELIMINARY; PRT; 1512 AA.  
 AC Q9WXJ5;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE GTF-S.  
 GN GTF.  
 OS Streptococcus criceti.  
 OG Plasmid pAM1.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 CC Streptococcus.  
 OX NCBI\_TaxID=1333;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HS-6;  
 RA Inoue M., Fukui K., Miyagi A.;  
 RT "S.cricetus glucosyltransferase(gtfs and gtf) genes.";  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB026123; BAA77237.1; -  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 14.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 DR Plasmid.  
 SQ SEQUENCE 1512 AA; 167145 MW; 4C03D9CBC601FC14 CRC64;

Query Match 70.0%; Score 77; DB 2; Length 1512;  
 Best Local Similarity 71.4%; Pred. No. 0.0031;  
 Matches 15; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Oy 2 PSYSFIRAHDSVDLIRNII 22  
 |||I:|||||:|  
 Db 560 PSYFVRAHDSVQTVAIQII 580

RESULT 15  
 Q9WXJ4 PRELIMINARY; PRT; 1338 AA.  
 AC Q9WXJ4;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE GTF-S.  
 GN GTF.  
 OS Streptococcus criceti.  
 OG Plasmid pAM1.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 CC Streptococcus.  
 OX NCBI\_TaxID=1333;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HS-6;  
 RA Inoue M., Fukui K., Miyagi A.;  
 RT "S.cricetus glucosyltransferase(gtfs and gtf) genes.";  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB026123; BAA77236.1; -  
 DR HSP; P06278; IVJS.  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 10.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 DR Plasmid.  
 KW SEQUENCE 1338 AA; 148558 MW; 0A90C98E10E15D99B CRC64;

Query Match 69.1%; Score 76; DB 2; Length 1338;  
 Best Local Similarity 72.7%; Pred. No. 0.0038;  
 Matches 16; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Oy 1 VPSYFIRAHDSVDLIRNII 22  
 |||I:|||||:|  
 Db 509 VPNYVFIHDSVQTRIAKII 530

Search completed: March 27, 2002, 14:26:08  
 Job time: 1680 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:57:58 ; Search time 198.55 Seconds  
(without alignments)  
8.208 Million cell updates/sec

Title: US-09-290-049a-16

Perfect score: 107

Sequence: 1 MANYIFIRAHDSVQTVIAKII 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_ll01.\*  
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	83.2	1577	17 AAR91047	Alpha-D-glucosyltr
2	64	59.8	1592	14 AAR32925	Glucosyltransferase
3	53	49.5	2057	21 AAB10667	L. mesenteroides a
4	44	41.1	282	21 AAG27318	Arabidopsis thalia
5	44	41.1	336	21 AAG49579	Arabidopsis thalia
6	44	41.1	423	21 AAG27317	Arabidopsis thalia
7	44	41.1	425	21 AAG27316	Arabidopsis thalia
8	44	41.1	477	21 AAG49578	Arabidopsis thalia
9	44	41.1	479	21 AAG49577	Arabidopsis thalia
10	44	41.1	583	22 AAE04782	Arabidopsis thalia
11	41.5	38.8	334	21 AAG12825	Arabidopsis thalia

12	40	37.4	349	22	AAB46747	R. marinus bacteri
13	40	37.4	871	22	AAB46727	R. marinus bacteri
14	40	37.4	2270	20	AAV24150	Candida albicans a
15	39.5	36.9	332	12	AAR13508	P. denitrificans CO
16	39	36.4	78	22	AAAM21090	Peptide #11326 enc
17	39	36.4	78	22	AAAM37289	Peptide #11326 enc
18	39	36.4	174	20	AAV19782	B. burgdorferi ant
19	39	36.4	176	22	AAG72801	Human olfactory re
20	39	36.4	182	20	AAV36750	PGL splice variant
21	39	36.4	185	20	AAV36753	PGL splice variant
22	39	36.4	195	20	AAV19781	B. burgdorferi ant
23	39	36.4	228	20	AAV36740	Truncated PGL prot
24	39	36.4	236	22	AAE04574	Human G-protein co
25	39	36.4	300	20	AAV36752	PGL splice variant
26	39	36.4	315	20	AAV36751	PGL splice variant
27	39	36.4	348	22	AAG71490	Human olfactory re
28	39	36.4	348	22	AAG72907	Human olfactory re
29	39	36.4	352	20	AAV36728	Human PGL protein
30	39	36.4	353	22	AAAM39591	Human polypeptide
31	39	36.4	353	22	AAB85532	Human secreted pro
32	39	36.4	353	22	AAB85552	Human secreted pro
33	39	36.4	353	22	AAB93595	Human protein sequ
34	39	36.4	364	20	AAV36729	Human PGL protein
35	39	36.4	364	22	AAU00665	Human lysophosphat
36	39	36.4	372	22	AAAM1377	Human polypeptide
37	39	36.4	680	21	AAV79196	Maize glutamine:fr
38	39	36.4	1091	10	AAAP98500	Partial sequence e
39	39	36.4	1600	21	AAG51095	Arabidopsis thalia
40	39	36.4	1608	21	AAG51094	Arabidopsis thalia
41	39	36.4	1625	21	AAG51093	Arabidopsis thalia
42	38	35.5	65	21	AAB18016	Eucalyptus grandis
43	38	35.5	257	20	AAV33674	R. sphaeroides 3-H
44	38	35.5	260	22	AAE59367	Human protein tyro
45	38	35.5	260	22	AAB59388	Human protein tyro

#### ALIGNMENTS

RESULT 1

AAR91047  
ID AAR91047 standard; Protein; 1577 AA.

XX AAR91047;

XX 22-MAY-1996 (first entry)

XX Alpha-D-glucosyltransferase.

XX Alpha-D-glucosyltransferase; primer-independent; soluble glucan;  
sucrose; transgenic plant; cloning; Escherichia coli;  
phage lambda-C13; vector; plasmid pGSG501; plasmid pGSG502;  
gene transfer; crop improvement; storage carbohydrate; pasture;  
feedstuff; senescence; dextran; binder; food; pharmaceutical.

XX Streptococcus salivarius strain ATCC 25975.

XX WO9606173-A1.

XX 29-FEB-1996.

XX 24-AUG-1995; 95WO-AU00527.

XX 24-AUG-1994; 94AU-0007643.

XX (GIFF/) GIFFARD P M.

XX (JACO/) JACQUES N A.

XX (SIMP/) SIMPSON C L.

XX Giffard PM, Jacques NA, Simpson CL;

XX WPI; 1996-151376/15.

XX N-PSDB; AAT13139.



Oy 3 NYIFRAHSDSEVQTVIK 20  
| | : | | | : | | |  
Db 759 nysfvrahdyaqdpirk 776

RESULT 4  
AAG27318  
ID AAG27318 standard; Protein; 282 AA.  
XX  
AC AAG27318;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 32106.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 23-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.

```
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 41.18; Score 44; DB 21; Length 282;
Best Local Similarity 50.08;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Oy 5 IFIRAHDSVOTVIK 20
Db 146 ifviphdsdvfsvar 161

RESULT 5
```

```
AAG49579
ID AAG49579 standard; Protein; 336 AA.
XX
AC AAG49579;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 62735.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
```

PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139550.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 09-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0148684.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.

PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 41.1%; Score 44; DB 21; Length 336;  
Best Local Similarity 50.0%; Pred. No. 9.1;

Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 5 IFTRAHDSVQTVIAK 20  
||: |||:|:|:|:  
Db 146 ifvlphdsdvfsvar 161

RESULT 6  
AAG27317  
ID AAG27317 standard; Protein; 423 AA.  
XX  
AC AAG27317;  
XX  
DT 17-OCT-2000 (first entry)  
XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 32105.  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX EPI033405-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.  
XX 09-MAR-1999; 99US-0123548.  
XX 23-MAR-1999; 99US-0125788.  
XX 25-MAR-1999; 99US-0126264.  
XX 29-MAR-1999; 99US-0126785.  
XX 01-APR-1999; 99US-0127462.  
XX 06-APR-1999; 99US-0128234.  
XX 08-APR-1999; 99US-0128714.  
XX 16-APR-1999; 99US-0129845.  
XX 19-APR-1999; 99US-0130077.  
XX 21-APR-1999; 99US-0130449.  
XX 23-APR-1999; 99US-0130510.  
XX 28-APR-1999; 99US-0130891.  
XX 30-APR-1999; 99US-0131449.  
XX 30-APR-1999; 99US-0132048.  
XX 04-MAY-1999; 99US-0132407.  
XX 05-MAY-1999; 99US-0132484.  
XX 05-MAY-1999; 99US-0132485.  
XX 06-MAY-1999; 99US-0132486.  
XX 07-MAY-1999; 99US-0132487.  
XX 11-MAY-1999; 99US-0132863.  
XX 14-MAY-1999; 99US-0134256.  
XX 14-MAY-1999; 99US-0134218.  
XX 14-MAY-1999; 99US-0134219.  
XX 14-MAY-1999; 99US-0134221.  
XX 14-MAY-1999; 99US-0134370.  
XX 18-MAY-1999; 99US-0134768.  
XX 19-MAY-1999; 99US-0134941.  
XX 20-MAY-1999; 99US-0135124.  
XX 21-MAY-1999; 99US-0135353.  
XX 24-MAY-1999; 99US-0135629.  
XX 25-MAY-1999; 99US-0136021.  
XX 27-MAY-1999; 99US-0136392.  
XX 28-MAY-1999; 99US-0136782.  
XX 01-JUN-1999; 99US-0137222.  
XX 03-JUN-1999; 99US-0137528.  
XX 04-JUN-1999; 99US-0137502.  
XX 07-JUN-1999; 99US-0137724.  
XX 08-JUN-1999; 99US-0138094.  
XX 10-JUN-1999; 99US-0138540.  
XX 10-JUN-1999; 99US-0138847.  
XX 14-JUN-1999; 99US-0139119.  
XX 16-JUN-1999; 99US-0139452.  
XX 16-JUN-1999; 99US-0139453.  
XX 17-JUN-1999; 99US-0139492.  
XX 18-JUN-1999; 99US-0139454.  
XX 18-JUN-1999; 99US-0139455.  
XX 18-JUN-1999; 99US-0139456.  
XX 18-JUN-1999; 99US-0139457.  
XX 18-JUN-1999; 99US-0139458.  
XX 18-JUN-1999; 99US-0139459.  
XX 18-JUN-1999; 99US-0139460.  
XX 18-JUN-1999; 99US-0139461.  
XX 18-JUN-1999; 99US-0139462.  
XX 18-JUN-1999; 99US-0139463.  
XX 18-JUN-1999; 99US-0139750.  
XX 18-JUN-1999; 99US-0139763.  
XX 21-JUN-1999; 99US-0139817.  
XX 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.

PR	30-AUG-1999;	99US-0151303.	XX	
PR	31-AUG-1999;	99US-0151438.	PN	EPI033405-A2.
PR	01-SEP-1999;	99US-0151930.	XX	
PR	07-SEP-1999;	99US-0152363.	PD	06-SEP-2000.
PR	10-SEP-1999;	99US-0153070.	XX	
PR	13-SEP-1999;	99US-0153758.	PF	25-FEB-2000; 2000EP-0301439.
PR	15-SEP-1999;	99US-0154018.	XX	
PR	16-SEP-1999;	99US-0154039.	XX	
PR	20-SEP-1999;	99US-0154779.	PR	25-FEB-1999;
PR	22-SEP-1999;	99US-0155139.	PR	05-MAR-1999;
PR	23-SEP-1999;	99US-0155486.	PR	09-MAR-1999;
PR	24-SEP-1999;	99US-0155659.	PR	23-MAR-1999;
PR	28-SEP-1999;	99US-0156458.	PR	25-MAR-1999;
PR	29-SEP-1999;	99US-0156596.	PR	99US-0126264.
PR	04-OCT-1999;	99US-0157117.	PR	99US-0126785.
PR	05-OCT-1999;	99US-0157153.	PR	01-APR-1999;
PR	06-OCT-1999;	99US-0157865.	PR	06-APR-1999;
PR	07-OCT-1999;	99US-0158029.	PR	08-APR-1999;
PR	08-OCT-1999;	99US-0158232.	PR	99US-0128714.
PR	12-OCT-1999;	99US-0158369.	PR	99US-0129845.
PR	13-OCT-1999;	99US-0159293.	PR	99US-0130077.
PR	13-OCT-1999;	99US-0159294.	PR	99US-0130449.
PR	13-OCT-1999;	99US-0159295.	PR	99US-0130510.
PR	14-OCT-1999;	99US-0159339.	PR	23-APR-1999;
PR	14-OCT-1999;	99US-0159331.	PR	28-APR-1999;
PR	14-OCT-1999;	99US-0159637.	PR	99US-0131449.
PR	14-OCT-1999;	99US-0159638.	PR	99US-0132048.
PR	18-OCT-1999;	99US-0159584.	PR	99US-0132407.
PR	21-OCT-1999;	99US-0160741.	PR	99US-0132484.
PR	21-OCT-1999;	99US-0160767.	PR	99US-0132486.
PR	21-OCT-1999;	99US-0160768.	PR	99US-0132487.
PR	21-OCT-1999;	99US-0160770.	PR	99US-0132863.
PR	21-OCT-1999;	99US-0160814.	PR	99US-0132863.
PR	21-OCT-1999;	99US-0160815.	PR	99US-0132863.
PR	22-OCT-1999;	99US-0160980.	PR	99US-0134256.
PR	22-OCT-1999;	99US-0160981.	PR	99US-0134218.
PR	22-OCT-1999;	99US-0160989.	PR	99US-0134219.
PR	25-OCT-1999;	99US-0161404.	PR	99US-0134221.
PR	25-OCT-1999;	99US-0161405.	PR	99US-0134370.
PR	25-OCT-1999;	99US-0161406.	PR	99US-0134768.
PR	26-OCT-1999;	99US-0161359.	PR	99US-0134941.
PR	26-OCT-1999;	99US-0161360.	PR	99US-0135124.
PR	26-OCT-1999;	99US-0161361.	PR	99US-0135353.
PR	28-OCT-1999;	99US-0161920.	PR	99US-0135629.
PR	28-OCT-1999;	99US-0161920.	PR	99US-0136021.
PR	28-OCT-1999;	99US-0161933.	PR	99US-0136392.
PR	29-OCT-1999;	99US-0162142.	PR	99US-0136782.
Query Match 41.1%; Score 44; DB 21; Length 423;			PR	99US-0138094.
Best Local Similarity 50.0%; Pred. No. 12;			PR	99US-0138540.
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;			PR	99US-0138847.
Qy	5 IFIRAHSEVQTVIAK 20		PR	99US-0139119.
Db	287 lfvlphdsdvfsvar 302		PR	99US-0139452.
RESULT 7			PR	99US-0139453.
ID	AAG27316		PR	99US-0139492.
AC	AAG27316 standard; Protein: 425 AA.		PR	99US-0139454.
XX			PR	99US-0139455.
XX			PR	99US-0139456.
XX			PR	99US-0139457.
DT	17-OCT-2000 (first entry)		PR	99US-0139458.
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 32104.		PR	99US-0139459.
KW	Protein identification; signal transduction pathway; metabolic pathway;		PR	99US-0139460.
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		PR	99US-0139461.
KW	termination sequence.		PR	99US-0139462.
XX			PR	99US-0139463.
OS	Arabidopsis thaliana.		PR	99US-0139750.

PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150584.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.

PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 41.1%; Score 44; DB 21; Length 425;  
Best Local Similarity 50.0%; Pred. No. 12;

Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 5 IFIRAHDSVQTVIAK 20

||: |||:|:|:

Db 289 ifvlphdsdvfsvlar 304

RESULT 8

AAG49578

ID AAG49578 standard; Protein; 477 AA.

XX AC AAG49578;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 62734.

XX KW Protein identification; signal transduction pathway; metabolic pathway;

XX KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EPI033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX XX



PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 31-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.

```
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 41.1%; Score 44; DB 21; Length 477;
Best Local Similarity 50.0%; Pred. No. 14;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 5 IFIRAHDSVQTVIAK 20
||: |||: ||: ||:
Db 287 ifvlphdsdvsfslr 302

RESULT 9
AAG49577
AAG49577 standard; Protein; 479 AA.
AC AAG49577;
XX XX
DT 18-OCT-2000 (first entry)
XX XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 62733.
XX XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX XX
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX XX
PD 06-SEP-2000.
XX XX
PF 25-FEB-2000; 2000EP-0301439.
XX XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.

99US-0128234.
99US-0128714.
99US-0129845.
99US-0130077.
99US-0130449.
99US-0130510.
99US-0130891.
99US-0131449.
99US-0132048.
99US-0132407.
99US-0132484.
99US-0132485.
99US-0132486.
99US-0132487.
99US-0132863.
99US-0134256.
99US-0134218.
99US-0134219.
99US-0134221.
99US-0134370.
99US-0134768.
99US-0134941.
99US-0135124.
99US-0135353.
99US-0135629.
99US-0136021.
99US-0136392.
99US-0136782.
99US-0137222.
99US-0137528.
99US-0137502.
99US-0137724.
99US-0138094.
99US-0138540.
99US-0138847.
99US-0139119.
99US-0139452.
99US-0139453.
99US-0139492.
99US-0139454.
99US-0139455.
99US-0139456.
99US-0139457.
99US-0139458.
99US-0139459.
99US-0139460.
99US-0139461.
99US-0139462.
99US-0139463.
99US-0139750.
99US-0139763.
99US-0139817.
99US-0139899.
99US-0140353.
99US-0140354.
99US-0140695.
99US-0140823.
99US-0140991.
99US-0141287.
99US-0141842.
99US-0142154.
99US-0142055.
99US-0142390.
99US-0142803.
99US-0142920.
99US-0142977.
99US-0143542.
99US-0143624.
99US-0144005.
99US-0144085.
99US-0144086.
99US-0144325.
99US-0144331.
```

```
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150366.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.

PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160788.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 41.1%; Score 44; DB 21; Length 479;
Best Local Similarity 50.0%; Pred. No. 14;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 5 IFIRAHDSVQTVIAK 20
   ||| ||| ||| |||
Db 289 IIVPHDSDFSVIAR 304

RESULT 10
AAE04782
ID AAE04782 standard; Protein; 583 AA.
AC AAE04782;
XX
XX 10-SEP-2001 (first entry)
XX
XX Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED1.
XX Neoxanthin cleavage enzyme; AtNCED1; abscisic acid; ABA; herbicide;
XX stress tolerance; transgenic plant; plant breeding; antisense-therapy;
XX plant growth protectant.
XX
XX Arabidopsis thaliana.
XX
XX EPI116794-A2.
XX
XX 18-JUL-2001.
XX
XX 11-JAN-2001; 2001EP-0300218.
XX
XX 13-JAN-2000; 2000JP-0010056.
XX
XX 11-JAN-2001; 2001JP-0003476.
XX
XX (RIKE ) RIKEN KK.
XX
XX Iuchi S, Kobayashi M, Shinozaki K;
XX WPI; 2001-400081/43.
XX
XX N-PSDB; AAD09394.
XX
XX A DNA encoding a protein with a neoxanthin cleavage activity for
XX producing transgenic plants with improved or decreased stress tolerance
XX
```

PS Claim 3; Fig 10; 101pp; English.

XX The invention relates to neoxanthin cleavage enzymes and their  
 CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key  
 CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.  
 CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a  
 CC plant when expressed in a plant cell. The invention also relates to  
 CC methods for increasing or decreasing stress tolerance in a plant by  
 CC introducing the DNA into the plant, and a transgenic plant into which a  
 CC neoxanthin cleavage enzyme is introduced. The improvement of stress  
 CC tolerance in plants is useful, for example in plant breeding. Neoxanthin  
 CC cleavage enzyme genes are useful for producing transgenic plants. An arid  
 CC land can be improved by growing transformant weed for several years and  
 CC then removing the weed by specifically lowering stress tolerance in the  
 CC weed by inducing an inducible promoter. The present sequence is  
 CC Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCEDI protein.  
 CC The AtNCEDI cDNA is obtained from an Arabidopsis plant-derived cDNA  
 CC library using a cDNA of the CPR65 (cowpea Responsive to Dehydration)  
 CC gene isolated from cowpea plant as a probe.

SQ Sequence 583 AA;

Query Match 41.1%; Score 44; DB 22; Length 583;  
 Best Local Similarity 47.4%; Pred. No. 18;  
 Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 4 YIFIRAHDSVQTVIAKII 22  
 ||| ||| :| :||  
 Db 530 yifchvdeetktselqii 548

RESULT 11

AAG12825  
 ID AAG12825 standard; Protein; 334 AA.

XX AC AAG12825;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 12087.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.

XX Arabidopsis thaliana.

EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126284.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 28-APR-1999; 99US-0130891.  
 PR 30-APR-1999; 99US-0131449.  
 PR 28-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139452.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.

PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 38.8%; Score 41.5; DB 21; Length 334;  
Best Local Similarity 23.3%; Pred. NO. 25;  
Matches 7; Conservative 11; Mismatches 1; Indels 11; Gaps 1;

QY 4 YIFIRAH-----DSEVQTVIAKII 22  
:||: :| |::|::|::|::|  
Db 27 flfvshlyvellffidskmtivsrli 56

RESULT 12  
AAB46747  
ID AAB46747 standard; Protein; 349 AA.  
XX  
AC AAB46747;  
XX  
XX  
DT 12-APR-2001 (first entry)  
XX  
DE R. marinus bacteriophage RM387 ORF056e protein.

XX  
KW Genome; thermophilic enzyme; washing powder; bleaching.  
OS Rhodothermus marinus.  
XX  
PN WO200075335-A2.  
XX  
PD 14-DEC-2000.  
XX  
PF 02-JUN-2000: 2000WO-IB00893.  
XX  
XX 02-JUN-1999. 99US-0137120.  
XX  
PA (DECO-) DECODE GENETICS EHF.

PI Hjorleifsgottir S, Hreggvidsson GO, Fridjonsson OH, Aevarsson A;  
PI Kristjansson JK;  
XX  
DR WPI; 2001-061727/07.  
DR N-PSDB; AAF25934.  
XX  
PT Isolated nucleic acid from bacteriophage RM 37P encoding enzymes useful  
PT in recombinant DNA technology.

XX  
PS Claim 20; Fig 7A-B; 42pp; English.

XX  
CC This invention describes a novel isolated nucleic molecule (I) comprising  
CC the genome of bacteriophage RM 378. The invention also describes (1) an  
CC isolated nucleic acid which encodes a polypeptide obtainable from  
CC bacteriophage RM 378, or its active derivative or fragment; (2) an  
CC isolated bacteriophage RM 378 (II); (3) a DNA construct (III) comprising  
CC operatively linked to a regulatory sequence; (4) a host cell comprising  
CC (III); and (5) an isolated polypeptide (IV) obtainable from (II), its  
CC active derivative or fragment. Bacteriophage RM 378 is useful for  
CC producing thermophilic enzymes useful in DNA research and commercial  
CC settings (e.g. proteases and lipases used in washing powder, hydrolytic  
CC enzymes used in bleaching). The isolated nucleic acid molecules and  
CC vectors are useful in the manufacture of encoded polypeptide, as probes  
CC for isolating homologous sequences (e.g. from other bacteriophage



Query Match 37.4%; Score 40; DB 20; Length 2270;  
 Best Local Similarity 52.9%; Pred. No. 4.7e+02;  
 Matches 9; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 6 FIRAHDESVQTVIAKII 22  
 I:III III:II:  
 Db 89 fvrahqg--htviskil 103

## RESULT 15

AAR13508  
 ID AAR13508 standard; Protein; 332 AA.

XX AC AAR13508;

XX DT 25-OCT-1991 (first entry)

XX DE P.denitrificans COB S.

cob gene; corrinoid; descobaltocorrinoid; cor gene.

OS Pseudomonas denitrificans.

XX WO91111518-A.

XX PD 08-AUG-1991.

XX PF 30-JAN-1991; 91WO-FR00054.

XX PR 31-JAN-1990; 90FR-0001137.

XX PA (RHON ) RHONE-POULENC BIOCH.

XX PI Blanche F, Meron B, Crouzet J, Debussche L, Levy-Schil S;

XX PI Thibaut D;

XX DR WPI; 1991-252650/34.

XX DR N-PSDB; AAQ13286.

XX PT New polypeptide(s) involved in cobalamin and cobamide

XX PT biosynthesis - and DNA encoding them, for amplification of

XX PT cobalamin, esp. coenzyme B12 prodn.

XX PS Claim 15; Fig 40; 299pp; French.

XX CC This sequence corresponds to one of 24 polypeptides obtained from

XX CC P.denitrificans and implicated in the biosynthesis of cobalamines

XX CC and/or cobamides, specifically in the conversion of precorrin-3 to

XX CC cobyrinic acid a,c-diamide. It is encoded by part of the 4.7kb

XX CC Sali-Sali-Sali-Sali-BgII fragment isolated from a

XX CC P.denitrificans genomic DNA bank constructed in vector pXL59.

XX CC See AAQ13284-Q13288.

XX SQ Sequence 332 AA;

Query Match 36.9%; Score 39.5; DB 12; Length 332;  
 Best Local Similarity 47.4%; Pred. No. 58;  
 Matches 9; Conservative 2; Mismatches 5; Indels 3; Gaps 1;

QY 3 NYFIRAHDESVQTVIAKI 21

II: III III III

Db 219 nyl---phdkevdivaakv 234

Search completed: March 27, 2002, 13:57:59  
 Job time: 527 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:27:05 ; Search time 53.4 seconds  
(without alignments)

14.419 Million cell updates/sec

Title: US-09-290-049a-14  
Perfect score: 119  
Sequence: 1 ANNHVSIVEAWSNDTPYLHD 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	119	100.0	1592	1 GTF2_STRDO	P27470 streptococc
2	119	100.0	1597	1 GTF1_STRDO	P11001 streptococc
3	108	90.8	1476	1 GTFB_STRMU	P08987 streptococc
4	99	83.2	1375	1 GTFB_STRMU	P13470 streptococc
5	74	62.2	1462	1 GTFD_STRMU	P49331 streptococc
6	67	56.3	1365	1 GTFD_STRDO	P29336 streptococc
7	49	41.2	336	1 UL16_EBV	P03221 epstein-bar
8	48	40.3	330	1 G3PC_LEIME	Q01558 leishmania
9	46	38.7	313	1 YM58_YEAST	Q03695 saccharomyc
10	45	37.8	337	1 G3P_MYCGE	P47543 mycoplasma
11	44	37.0	1122	1 YG3C_YEAST	P53280 saccharomyc
12	44	37.0	1131	1 YPO4_CAEEL	P33015 caenorhabdi
13	44	37.0	3358	1 PGCY_MOUSE	Q62059 mus muscul
14	43.5	36.6	491	1 G6PD_ECOLI	P22992 escherichia
15	43	36.1	275	1 YG33_YEAST	P53271 saccharomyc
16	43	36.1	330	1 G3PC_TRYBB	P10097 trypanosoma
17	43	36.1	346	1 Y943_MYCTU	P171566 mycobacteri
18	43	36.1	441	1 HGD_STRCO	Q9S2B5 streptomyce
19	43	36.1	603	1 DOPO_HUMAN	P09172 homo sapien
20	43	36.1	899	1 SUHW_DROVI	Q08876 drosophila
21	42.5	35.7	470	1 AMP2_STRCO	Q60394 streptomyce
22	42.5	35.7	491	1 G6PD_ERWCH	P37986 erwinia chr
23	42	35.3	279	1 CG16_YEAST	P24867 saccharomyc
24	42	35.3	334	1 G3P_BACST	P00362 bacillus st
25	42	35.3	335	1 G3P_BACCO	P15115 bacillus co
26	42	35.3	432	1 HGD_PSEAE	Q9X4G0 pseudomonas
27	42	35.3	775	1 AD28_HUMAN	Q9UKQ2 homo sapien
28	42	35.3	776	1 AD28_MACFA	Q9XSL6 macaca fasc
29	42	35.3	1627	1 ADP1_MYCPN	P11311 mycoplasma
30	42	35.3	3898	1 POLG_HCVB	P21530 hog cholera
31	41.5	34.9	416	1 HGD_LGPN	Q9S4T0 legionella
32	41.5	34.9	468	1 KG3H_DICDI	P51136 dictyosteli
33	41.5	34.9	578	1 VAC8_YEAST	P39968 saccharomyc

34	41	34.5	130	1	UCR6_SCHMA	Q01374 schistosoma
35	41	34.5	236	1	UCR4_TOBAC	P51134 nicotiana t
36	41	34.5	241	1	HFB1_HAEIN	P35757 haemophilus
37	41	34.5	241	1	HFB2_HAEIN	P45991 haemophilus
38	41	34.5	301	1	Y186_MYCPN	P75265 mycoplasma
39	41	34.5	383	1	DUHM_BPSP1	P13654 bacterioph
40	41	34.5	621	1	DCTB_RHIME	P13633 rhizobium m
41	41	34.5	1139	1	INAL_CAEEL	Q03600 caenorhabdi
42	41	34.5	1634	1	DPOL_METJA	O58295 methanococc
43	41	34.5	2485	1	PTND_HUMAN	Q12923 homo sapien
44	41	34.5	3341	1	POLG_MCPA	P35515 m genome po
45	41	34.5	3898	1	POLG_BVDVS	Q01499 bovine vira

## ALIGNMENTS

RESULT 1						
GTF2_STRDO	ID	GTF2_STRDO	STANDARD;	PRT;	1592 AA.	
AC	P27470:					
DT	01-AUG-1992 (Rel. 23, Created)					
DT	01-AUG-1992 (Rel. 23, Last sequence update)					
DT	15-DEC-1998 (Rel. 37, Last annotation update)					
DE	GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)					
DE	(SUCROSE 6-GLUCOSYLTRANSFERASE).					
OS	Streptococcus downei (Streptococcus sobrinus).					
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;					
OC	Streptococcus.					
OX	NCBI_TaxID=1317;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=6715;					
RC	MEDLINE=01113227; PubMed=1704006;					
RA	Ado H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,					
RA	Kagawa H.;					
RT	"Peptide sequences for sucrose splitting and glucan binding within					
RT	Streptococcus sobrinus glucosyltransferase (water-insoluble glucan					
RT	synthetase)";					
RL	J. Bacteriol. 173:989-996(1991).					
CC	-!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT					
CC	TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE					
CC	OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE					
CC	AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.					
CC	-!- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -					
CC	D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).					
CC	-!- SUBCELLULAR LOCATION: SECRETED.					
CC	-!- DISEASE: DENTAL CARIES.					
CC	-!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA					
CC	1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES					
CC	WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH					
CC	FORMS OF GLUCANS.					
CC	-!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-					
CC	BINDING PROTEIN FROM S. MUTANS.					
CC	-----					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/					
CC	or send an email to license@isb-sib.ch).					
CC	-----					
CC	EMBL; D90213; BAA14241.1;					
DR	PIR; A38175; A38175.					
DR	HSP; P00695; 2HEE.					
DR	InterPro: IPR002479; CW binding.					
DR	InterPro: IPR003318; Glyco_hydro_70.					
DR	Pfam; PF01473; CW_binding_1; 16.					
DR	Pfam; PF02324; Glyco_hydro_70; 1.					
KW	Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.					
FT	SIGNAL 1 38					
FT	POTENTIAL.					
FT	CHAIN 39 1592					
FT	GLUCOSYLTRANSFERASE-I.					

FT DOMAIN 39 1044 CATALYTIC (APPROXIMATE).  
 FT DOMAIN 1093 1592 GLUCAN-BINDING (APPROXIMATE).  
 FT DOMAIN 1093 1592 6.5 X TANDEM REPEATS.  
 FT REPEAT 1093 1142 1.  
 FT REPEAT 1158 1207 2.  
 FT REPEAT 1222 1272 3.  
 FT REPEAT 1287 1337 4.  
 FT REPEAT 1402 1451 5.  
 FT REPEAT 1514 1563 6.  
 FT REPEAT 1577 1592 7 (INCOMPLETE).  
 SQ SEQUENCE 1592 AA; 176167 MW; BCOA66D079351ECF CRC64;

Query Match 100.0%; Score 119; DB 1; Length 1592;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-10; Indels 0; Gaps 0;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ANNHVSIVEAWSNDTPTLHD 21  
 |||||  
 477 ANNHVSIVEAWSNDTPTLHD 497

## RESULT 2

GTFL\_STRDO ID GTFL\_STRDO STANDARD; PRT: 1597 AA.  
 AC P11001;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)  
 DE (SUCCROSE 6-GLUCOSYLTRANSFERASE).  
 GN GTFI.  
 OS Streptococcus downei (Streptococcus sobrinus).  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 ON NCBI\_TaxID=1317;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MFE28;  
 RX MEDLINE=87308014; PubMed=3040686;  
 RA Ferretti J.J., Gilpin M.L., Russell R.R.B.;  
 RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus  
 sobrinus MFE28";  
 RL J. Bacteriol. 169:4271-4278(1987).  
 CC -|- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
 AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -|- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =  
 D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -|- SUBCELLULAR LOCATION: SECRETED.  
 CC -|- DISEASE: DENTAL CARIES.  
 CC -|- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
 WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
 FORMS OF GLUCANS.  
 CC -|- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
 BINDING PROTEIN FROM S. MUTANS.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; M17391; AAC63063.1;  
 InterPro; IPR002479; CW binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 19.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.

FT SIGNAL 1 38 POTENTIAL.  
 FT CHAIN 39 1597 GLUCOSYLTRANSFERASE-I.  
 FT DOMAIN 39 1050 CATALYTIC (APPROXIMATE).  
 FT DOMAIN 1099 1597 GLUCAN-BINDING (APPROXIMATE).  
 FT DOMAIN 1099 1597 1.25 A, 2 B, AND 5 AC REPEATS.  
 FT REPEAT 1099 1132 A REPEAT.  
 FT REPEAT 1163 1213 AC REPEAT.  
 FT REPEAT 1227 1277 AC REPEAT.  
 FT REPEAT 1292 1342 AC REPEAT.  
 FT REPEAT 1352 1399 B REPEAT.  
 FT REPEAT 1406 1455 AC REPEAT.  
 FT REPEAT 1465 1512 B REPEAT.  
 FT REPEAT 1519 1568 AC REPEAT.  
 FT REPEAT 1582 1597 A REPEAT (INCOMPLETE).  
 SQ SEQUENCE 1597 AA; 177080 MW; B9E86A200868798E CRC64;

Query Match 100.0%; Score 119; DB 1; Length 1597;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-10; Indels 0; Gaps 0;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANNHVSIVEAWSNDTPTLHD 21  
 |||||  
 DB 483 ANNHVSIVEAWSNDTPTLHD 503

## RESULT 3

GTFL\_STRMU ID GTFL\_STRMU STANDARD; PRT: 1476 AA.  
 AC P08987; O69381; O69384; O69387; O69390; O69396;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)  
 DE (SUCCROSE 6-GLUCOSYLTRANSFERASE).  
 GN GTFB.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 ON NCBI\_TaxID=1309;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GS-5;  
 RX MEDLINE=87308013; PubMed=3040685;  
 RA Shiroza T., Ueda S., Kuramitsu H.K.;  
 RT "Sequence analysis of the gtfB gene from Streptococcus mutans";  
 RL J. Bacteriol. 169:4263-4270(1987).  
 CC -|-  
 CC SEQUENCE FROM N.A.  
 CC STRAIN=MT4239, MT4245, MT4251, MT4467, AND MT8148;  
 CC MEDLINE=98231643; PubMed=9570124;  
 CC Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
 CC Kimura S., Hamada S.;  
 CC "Molecular analyses of glucosyltransferase genes among strains of  
 CC Streptococcus mutans";  
 CC FEMS Microbiol. Lett. 161:331-336(1998).  
 CC -|- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
 AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -|- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =  
 D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -|- SUBCELLULAR LOCATION: SECRETED.  
 CC -|- DISEASE: DENTAL CARIES.  
 CC -|- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
 WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
 FORMS OF GLUCANS.  
 CC -|- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
 BINDING PROTEIN FROM S. MUTANS.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; M17391; AAC63063.1;  
 InterPro; IPR002479; CW binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 19.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC EMBL; M17361; AAA88588.1; -
CC EMBL; D88651; BAA26101.1; -
CC EMBL; D88654; BAA26105.1; -
CC EMBL; D88657; BAA26109.1; -
CC EMBL; D88660; BAA26113.1; -
CC EMBL; D88977; BAA26119.1; -
CC PIR; B33135; B33135.
CC InterPro; IPR002479; CW_binding.
CC InterPro; IPR003318; Glyco_hydro_70.
CC Pfam; PF01473; Glyco_binding_1; 13.
CC Pfam; PF02324; Glyco_hydro_70; 1.
CC Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.
CC SIGNAL
CC CHAIN 1 34
CC DOMAIN 35 1476
CC REPEAT 1097 1476
CC REPEAT 1161 1470
CC REPEAT 1161 1210
CC REPEAT 1225 1275
CC REPEAT 1290 1340
CC REPEAT 1355 1405
CC REPEAT 1420 1470
CC VARIANT 62 62
CC VARIANT 65 65
CC VARIANT 68 68
CC VARIANT 78 78
CC VARIANT 86 86
CC VARIANT 89 89
CC VARIANT 168 168
CC VARIANT 276 276
CC VARIANT 399 399
CC VARIANT 474 474
CC VARIANT 512 512
CC VARIANT 519 519
CC VARIANT 701 701
CC VARIANT 708 708
CC VARIANT 938 938
CC VARIANT 952 957
CC VARIANT 963 964
CC VARIANT 968 970
CC VARIANT 1086 1086
CC VARIANT 1158 1158
CC VARIANT 1163 1163
CC VARIANT 1168 1168
CC VARIANT 1182 1182
CC VARIANT 1234 1234
CC VARIANT 1263 1263
CC VARIANT 1263 1263
CC VARIANT 1264 1264
CC VARIANT 1272 1272
CC VARIANT 1329 1329
CC VARIANT 1394 1394
CC VARIANT 1402 1402
CC VARIANT 1459 1459
CC CONFLICT 570 570
CC CONFLICT 800 817
CC CONFLICT 1310 1310
CC CONFLICT 1476 AA; 165685 MW; 3479BG2B07694D98 CRC64;

```

Query Match 90.8%; Score 108; DB 1; Length 1476;  
 Best Local Similarity 85.7%; Pred. No. 5.5e-09;  
 Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANNHVIVEAWSNDNDTPYLHD 21  
 DB 481 ANDHLSLEAWSNDNDTPYLHD 501

## RESULT 4

```

GTFC_STRMU STANDARD; PRT; 1375 AA.
AC P13470; P05427;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI PRECURSOR (EC 2.4.1.5) (GTF-SI)
DE (DEXTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTFC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=89137980; PubMed=2976010;
RA Ueda S., Shiroza T., Kuramitsu H.K.;
RT "Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.";
RL Gene 69:101-109(1988).
RN [2]
RP SEQUENCE OF 1-349 FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N+1) =
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DISEASE: DENTAL CARIES.
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE) GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S. MUTANS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

EMBL; M22054; AAA88592.1; -  
 EMBL; M17361; AAA88589.1; -  
 PIR; J0345; J0345.  
 PIR; C33135; C33135.  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 7.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 DR Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.  
 FT SIGNAL 1 34  
 FT CHAIN 35 1375 GLUCOSYLTRANSFERASE-SI.

FT DOMAIN 35 1050 CATALYTIC (APPROXIMATE).  
FT DOMAIN 1126 1375 GLUCAN-BINDING (APPROXIMATE).  
FT DOMAIN 1126 1375 2.4 A, 1 C AND 1 AC REPEATS.  
FT REPEAT 1126 1159 A REPEAT.  
FT REPEAT 1169 1200 A REPEAT.  
FT REPEAT 1227 1238 C REPEAT.  
FT REPEAT 1253 1303 A REPEAT.  
FT REPEAT 1318 1330 A REPEAT (INCOMPLETE).  
SQ SEQUENCE 1375 AA; 153022 MW; D4B80CBEE0AAE13 CRC64;

Query Match 83.28; Score 99; DB 1; Length 1375;  
Best Local Similarity 81.08; Pred. No. 1.3e-07;  
Matches 17; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ANNHVSIVEAWSNDTPYLHD 21  
507 ANDHLSILEAWSNDTPYLHD 527  
11:11:1111111111111111

RESULT 5  
GTFD\_STRMU STANDARD; PRT; 1462 AA.  
AC P49331; O69386; O69389; O69392; O69398;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)  
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
GN GTFD.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID:1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-GS-5;  
RX MEDLINE=91100958; PubMed=2148600;  
RA Honda O., Kato C., Kuramitsu H.K.;  
RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding  
the glucosyltransferase-S enzyme."  
RL J. Gen. Microbiol. 136:2099-2105(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MT4239, MT4245, MT4251, MT4467, AND MT8148;  
RC MEDLINE=98231643; PubMed=9570124;  
RW Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ocshima T., Sobue S.,  
Kimura S., Hamada S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of  
Streptococcus mutans."  
RL FEMS Microbiol. Lett. 161:331-336(1998).  
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =  
D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
CC -1- SUBCELLULAR LOCATION: SECRETED.  
CC -1- DISEASE: DENTAL CARIES.  
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
FORMS OF GLUCANS.  
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
BINDING PROTEIN FROM S. MUTANS.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC DR EMBL; M29296; AAA26895.1; -  
DR EMBL; D88653; BAA26103.1; -  
DR EMBL; D88656; BAA26107.1; -  
DR EMBL; D88659; BAA26111.1; -  
DR EMBL; D88662; BAA26115.1; -  
DR EMBL; D89979; BAA26121.1; -  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 1.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.  
FT SIGNAL 1 ?  
FT CHAIN ? 1462  
FT DOMAIN 1232 1423  
FT REPEAT 1232 1295  
FT REPEAT 1296 1359  
FT REPEAT 1360 1423  
FT VARIANT 58 58  
FT VARIANT 68 68  
FT VARIANT 81 81  
FT VARIANT 113 113  
FT VARIANT 122 122  
FT VARIANT 132 132  
FT VARIANT 135 135  
FT VARIANT 202 202  
FT VARIANT 255 255  
FT VARIANT 275 275  
FT VARIANT 288 288  
FT VARIANT 301 301  
FT VARIANT 313 313  
FT VARIANT 317 317  
FT VARIANT 328 328  
FT VARIANT 350 350  
FT VARIANT 628 633  
FT VARIANT 688 688  
FT VARIANT 726 732  
FT VARIANT 726 730  
FT VARIANT 964 964  
FT VARIANT 1019 1019  
FT VARIANT 1059 1060  
FT VARIANT 1060 1060  
FT VARIANT 1080 1080  
FT VARIANT 1142 1142  
FT VARIANT 1198 1198  
FT VARIANT 1220 1220  
FT VARIANT 1280 1280  
FT VARIANT 1282 1282  
FT VARIANT 1290 1290  
FT VARIANT 1311 1311  
FT VARIANT 1403 1403  
FT VARIANT 1425 1425  
FT VARIANT 1449 1449  
FT CONFLICT 1428 1462  
FT SEQUENCE 1462 AA; 163512 MW; 5C6541F0DCB0DF00 CRC64;

-----  
K -> E (IN STRAIN MT4467).  
A -> S (IN STRAINS MT4239 AND MT4245).  
A -> T (IN STRAINS MT4251 AND MT8148).  
T -> I (IN STRAINS MT4239 AND MT4245).  
A -> V (IN STRAINS MT4239, MT4245 AND  
MT8148).  
S -> A (IN STRAINS MT4239, MT4245, MT4251  
AND MT8148).  
A -> V (IN STRAIN MT4245).  
V -> L (IN STRAIN MT4239).  
D -> N (IN STRAIN MT8148).  
E -> D (IN STRAINS MT4239, MT4245 AND  
MT4251).  
D -> N (IN STRAINS MT4239, MT4245 AND  
MT4251).  
Q -> H (IN STRAIN MT4245).  
D -> N (IN STRAINS MT4239 AND MT4251).  
E -> K (IN STRAIN MT4239).  
V -> F (IN STRAIN MT4239).  
F -> L (IN STRAINS MT4239, MT4251 AND  
MT4467).  
KKRYTQ -> EKEYTL (IN STRAIN MT4251).  
A -> S (IN STRAIN MT4239).  
TDQGEA -> ADKNGS (IN STRAIN MT4251).  
TDQGS -> ADKGN (IN STRAINS MT4239 AND  
MT4245).  
D -> Y (IN STRAIN MT4251).  
E -> K (IN STRAINS MT4245 AND MT4251).  
LG -> IR (IN STRAIN MT4251).  
G -> R (IN STRAIN MT4245).  
G -> R (IN STRAIN MT4239).  
Q -> H (IN STRAINS MT4239, MT4245,  
MT4251, MT4467 AND MT8148).  
S -> N (IN STRAIN MT4239).  
Y -> C (IN STRAINS MT4251 AND MT4467).  
F -> L (IN STRAIN MT4467).  
Q -> P (IN STRAIN MT4245).  
K -> T (IN STRAIN MT4245).  
N -> D (IN STRAIN MT4245).  
D -> G (IN STRAINS MT4239, MT4245, MT4251  
AND MT8148).  
R -> G (IN STRAINS MT4239, MT4245,  
MT4251, MT4467 AND MT8148).  
R -> K (IN STRAIN MT4467).  
RRYDKNSGNVKNVYVTLANGRRIGIDRWGIARY ->  
VYR (IN REF. 1).

Query Match 62.28; Score 74; DB 1; Length 1462;  
Best Local Similarity 66.78; Pred. No. 0.0013;  
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 ANNHVSIVEAWSNDTPYLHD 21  
11:11:1111111111111111

Db 495 AINHLSILEAWSNDNDPQYNKD 515

```
RESULT 6
GTF5_STRDO
ID GTF5_STRDO STANDARD; PRT; 1365 AA.
AC P29336;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTF5.
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1317;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MFE28;
RX MEDLINE=90316665; PubMed=2142479;
"Analysis of the Streptococcus downei gtf5 gene, which specifies a
glucosyltransferase that synthesizes soluble glucans.";
RI FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -! CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -
D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).
CC -! ENZYME REGULATION: GLUCAN SYNTHESIS BY GTF-S IS INDEPENDENT OF
PRIMER GLUCAN UNLIKE GTF-I.
CC -! DISEASE: DENTAL CARIES.
CC -! MISCELLANEOUS: GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA
1,6-GLUCOSE).
CC -! SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
BINDING PROTEIN FROM S. MUTANS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M30943; AAA26898.1; -.
DR PIR; A41483; A41483.
DR InterPro; IPR002479; CW_binding.
DR Pfam; PF01473; CW_binding.1; 10.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1 36
FT CHAIN 37 1365
FT DOMAIN 37 1050
FT DOMAIN 1083 1365
FT DOMAIN 1083 1365
FT REPEAT 1083 1131
FT REPEAT 1150 1199
FT REPEAT 1225 1274
FT REPEAT 1289 1339
FT REPEAT 1353 1365
SQ SEQUENCE 1365 AA; 151590 MW; 167296B5A2E8C476 CRC64;
```

Query Match 56.38; Score 67; DB 1; Length 1365;  
Best Local Similarity 57.18; Pred. No. 0.016;  
Matches 12; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Oy 1 ANNHVSIVEAWSNDNDPTPLHD 21  
| : : : : : | | | | |

Db 467 AIDHLSILEAWSNDNDYVKD 487

```
RESULT 7
UL16_EBV
ID UL16_EBV STANDARD; PRT; 336 AA.
AC P03221;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE BGLF2 PROTEIN.
DE BGLF2.
GN Epstein-barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84270667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
Tufnell P.S., Barrell B.G.;
RA "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RT Nature 310:207-211(1984).
RL Nature 310:207-211(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92113548; PubMed=1662696;
Chen M.R., Hsu T.Y., Lin S.W., Chen J.Y., Yang C.S.;
"Cloning and characterization of cDNA clones corresponding to
transcripts from the BamHI G region of the Epstein-Barr virus genome
and expression of BGLF2.";
RT J. Gen. Virol. 72:3047-3055(1991).
CC -! SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL16,
HSV-6 ORF11, EBV-1 46, HCMV UL94, EBV BDLF2, AND VZV 44.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M60514; AAA45871.1; -.
DR EMBL; V01555; CAA24831.1; -.
DR EMBL; S77132; AAB21113.1; -.
DR PIR; A03784; QQBE40.
DR PIR; S33036; S33036.
KW Late protein.
SQ SEQUENCE 336 AA; 36888 MW; 840937A416D5584C CRC64;
```

Query Match 41.28; Score 49; DB 1; Length 336;  
Best Local Similarity 41.28; Pred. No. 2.1;  
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Oy 1 ANNHVSIVEAWSNDNDTP 17  
| : : : : : |  
Db 216 AGAHWNILRGWTEDDSP 232

```
RESULT 8
G3PC_LEIME
ID G3PC_LEIME STANDARD; PRT; 330 AA.
AC Q01558;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE, CYTOSOLIC (EC 1.2.1.12)
DE (GAPDH).
GN GAPC.
```

OS Leishmania mexicana.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.

```

OX NCBI_TaxID=5665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SSP. MEXICANA;
RX MEDLINE=93063042; PubMed=1435864;
RA Hannaert V., Blaauw M., Kohl L., Allert S., Opperdoes F.R.,
RA Michels P.A.M.;
RT "Molecular analysis of the cytosolic and glycosomal glyceraldehyde-3-
RT phosphate dehydrogenase in Leishmania mexicana.";
RL Mol. Biochem. Parasitol. 55:115-126(1992);
CC -!- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE
CC + NAD(+) = 1,3-DIPHOSPHATEGLYCERATE + NADH.
CC -!- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -!- SUBUNIT: HOMOTETRAMER.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE
CC DEHYDROGENASE FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; X65220; CAA46323.1; .
DR PIR; S25142; S25142.
DR PIR; B48445; B48445.
DR HSSP; P06977; 1GAD.
DR InterPro; IPR000173; GAP_DH.
DR Pfam; PF00044; gpdh; 1.
DR PRINTS; PS00078; G3PDHDEGNASE.
DR PROSITE; PS00071; GAPDH; 1.
KW Glycolysis; Oxidoreductase; NAD.
FT INIT_MET 0 0 BY SIMILARITY.
FT BINDING 148 148 GLYCERALDEHYDE 3-PHOSPHATE.
FT ACT_SITE 175 175 ACTIVATES THIOL GROUP DURING CATALYSIS.
SQ SEQUENCE 330 AA; 35511 MW; ED4B6D8E8A207F1E CRC64;

Query Match 40.3%; Score 48; DB 1; Length 330;
Best Local Similarity 42.1%; Pred. No. 3;
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

2 NNHVSIVEAWSNDTPYLH 20
1:1 :|11111
300 NDHFVKLVSYDNETGYSH 318

RESULT 9
YH58_YEAST
ID YH58_YEAST STANDARD; PRT; 313 AA.
AC Q03695;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHEICAL 35.0 KDA PROTEIN IN PFK2-HEP1 INTERGENIC REGION.
GN YMR206W OR YH8325.07.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Odell C., Bowman S., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: SOME, TO YEAST YMR014W.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

```

```

use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; Z48755; CAA8648.1; .
DR SGD; S0004819; YMR206W.
KW Hypothetical protein.
FT DOMAIN 3 POLY-SER.
FT DOMAIN 146 149 POLY-GLN.
FT DOMAIN 246 252 POLY-SER.
SQ SEQUENCE 313 AA; 35017 MW; 9D92BFDE982577F0 CRC64;

Query Match 38.7%; Score 46; DB 1; Length 313;
Best Local Similarity 47.4%; Pred. No. 5.8;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Oy 2 NNHVSIVEAWSNDTPYLH 20
111 :|11111
Db 125 NNHQFIARRFSDSHIPSLH 143

RESULT 10
G3P_MYCGE
ID G3P_MYCGE STANDARD; PRT; 337 AA.
AC P47543;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (EC 1.2.1.12) (GAPDH).
GN GAPA OR GAP OR MG301.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 1-81 AND 279-337 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
CC -!- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE
CC + NAD(+) = 1,3-DIPHOSPHATEGLYCERATE + NADH.
CC -!- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE
CC DEHYDROGENASE FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; U39710; AAC71523.1; .

```



J. Biol. Chem. 270:3914-3918(1995).

-1- FUNCTION: MAY PLAY A ROLE IN INTERCELLULAR SIGNALING AND IN CONNECTING CELLS WITH THE EXTRACELLULAR MATRIX. MAY TAKE PART IN THE REGULATION OF CELL MOTILITY, GROWTH AND DIFFERENTIATION. BINDS HYALURONIC ACID.

-1- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX.

-1- ALTERNATIVE PRODUCTS: AT LEAST 4 ISOFORMS; V0 (SHOWN HERE), V1, V2 AND V3: ARE PRODUCED BY ALTERNATIVE SPLICING.

-1- TISSUE SPECIFICITY: V2 IS FOUND ONLY IN BRAIN.

-1- DEVELOPMENTAL STAGE: DISAPPEARS AFTER THE CARTILAGE DEVELOPMENT.

-1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.

-1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.

-1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.

-1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.

-1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; D16263; BAA03796.1; -.

EMBL; D28599; -; NOT\_ANNOTATED\_CDS.

EMBL; D32040; BAA06802.1; -.

HSSP; P00740; IIXA.

MGD; MGI:102889; Cspg2.

InterPro; IPR000152; Asx\_hydroxyl.

InterPro; IPR000561; EGF-like.

InterPro; IPR000742; EGF\_2.

InterPro; IPR001881; EGF\_Ca.

InterPro; IPR001438; EGF\_II.

InterPro; IPR003599; Ig.

InterPro; IPR003006; Ig\_MHC.

InterPro; IPR000538; Link.

InterPro; IPR000436; Sushi\_SCR\_CCP.

InterPro; IPR001304; lectin\_c.

Pfam; PF00008; EGF; 2.

Pfam; PF00047; Ig; 1.

Pfam; PF00059; lectin\_c; 1.

Pfam; PF00084; sushi; 1.

Pfam; PF00193; Xlink; 2.

PRINTS; PR00010; EGF\_BLOOD.

ProDom; PD000918; Link; 2.

SMART; SM00032; CCP; 1.

SMART; SM00034; CLECT; 1.

SMART; SM00179; EGF\_CA; 1.

SMART; SM00001; EGF\_like; 1.

SMART; SM00409; IG; 1.

SMART; SM00445; LINK; 2.

PROSITE; PS00010; ASX\_HYDROXYL; 1.

PROSITE; PS00022; EGF\_1; 2.

PROSITE; PS01186; EGF\_2; 1.

PROSITE; PS01187; EGF\_CA; 1.

PROSITE; PS01241; LINK; 2.

PROSITE; PS00615; C\_TYPE\_LECTIN\_1; 1.

PROSITE; PS50041; C\_TYPE\_LECTIN\_2; 1.

GLYCOPROTEIN; Proteoglycan; Lectin; Extracellular matrix: Sushi; Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain; Hyaluronic acid; Alternative splicing.

KW SIGNAL 1 20 POTENTIAL.

FT CHAIN 21 3358

FT DOMAIN 37 137

FT DOMAIN 167 244

FT DOMAIN 265 334

FT DOMAIN 348 1308

FT DOMAIN 1309 3052

FT DOMAIN 3052 3088

FT DOMAIN 3090 3126

EGF-LIKE 1.

EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).

FT DOMAIN 3129 3255 C-TYPE LECTIN.

FT DOMAIN 3258 3316 SUSHI.

FT DISULFID 44 130 BY SIMILARITY.

FT DISULFID 172 243 BY SIMILARITY.

FT DISULFID 196 217 BY SIMILARITY.

FT DISULFID 270 333 BY SIMILARITY.

FT DISULFID 294 315 BY SIMILARITY.

FT DISULFID 3056 3067 BY SIMILARITY.

FT DISULFID 3061 3076 BY SIMILARITY.

FT DISULFID 3078 3087 BY SIMILARITY.

FT DISULFID 3094 3105 BY SIMILARITY.

FT DISULFID 3099 3114 BY SIMILARITY.

FT DISULFID 3116 3125 BY SIMILARITY.

FT DISULFID 3132 3143 BY SIMILARITY.

FT DISULFID 3160 3252 BY SIMILARITY.

FT DISULFID 3228 3244 BY SIMILARITY.

FT DISULFID 3259 3302 BY SIMILARITY.

FT DISULFID 3288 3315 BY SIMILARITY.

FT CARBOHYD 57 57 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 330 330 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 351 351 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 441 441 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 807 807 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 914 914 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 951 951 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1305 1305 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1372 1372 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1679 1679 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 2054 2054 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 2244 2244 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 2362 2362 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 2627 2627 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 3030 3030 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 3332 3332 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 3342 3342 N-LINKED (GLCNAC. .) (POTENTIAL).

FT VARSPLIC 349 1308 MISSING (IN ISOFORM V1).

FT VARSPLIC 1309 3052 MISSING (IN ISOFORM V2).

FT VARSPLIC 349 3052 MISSING (IN ISOFORM V3).

FT CONFLICT 348 348 P -> R (IN REF. 2).

SEQ SEQUENCE 3358 AA; 366938 MW; 071B80026BC0762D CRC64;

Query Match 37.0%; Score 44; DB 1; Length 3358;

Best Local Similarity 50.0%; Pred. No. 2e+02;

Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 3 NHVSIVEAWSNDTPY 18

Db 510 NHISLKELPEKNKTPY 525

RESULT 14

G6PD\_ECOLI STANDARD; PRT; 491 AA.

ID G6PD\_ECOLI AC P22992; P78069; Q60134; Q60139;

DT 01-AUG-1991 (Rel. 19, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DE 20-AUG-2001 (Rel. 40, Last annotation update)

GN ZWF OR B1852.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OC NCBI\_TaxID=562;

OX [1]

RN SEQUENCE FROM N.A.

RC STRAIN=K12;

RX MEDLINE=91123224; PubMed=1704005;

RA Rowley D.L., Wolf R.E. Jr.;

RT "Molecular characterization of the Escherichia coli K-12 zwf gene encoding glucose 6-phosphate dehydrogenase.";

RL J. Bacteriol. 173:968-977(1991).

RN [2]



RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of *Escherichia coli* K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97251358; PubMed=9097040;  
 RA Icho T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,  
 RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,  
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,  
 RA Saito N., Sampei G., Seki Y., Sivasubram S., Tagami H.,  
 RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;  
 RT "A 460-kb DNA sequence of the *Escherichia coli* K-12 genome  
 corresponding to the 40.1-50.0 min region on the linkage map.";  
 RL DNA Res. 3:379-392(1996).  
 RN [4]  
 RP SEQUENCE OF 72-368 FROM N.A.  
 RC STRAIN=VARIOUS ECOR STRAINS;  
 RX MEDLINE=95064015; PubMed=7973728;  
 RA Guttman D.S., Dykhuizen D.E.;  
 RT "Clonal divergence in *Escherichia coli* as a result of recombination,  
 not mutation.";  
 RL Science 266:1380-1383(1994).  
 RN [5]  
 RP SEQUENCE OF 321-491 FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=93345818; PubMed=8344525;  
 RA Carter A.T., Pearson B.M., Dickinson J.R., Lancashire W.E.;  
 RT "Sequence of the *Escherichia coli* K-12 edd and eda genes of the  
 Entner-Doudoroff pathway.";  
 RL Gene 130:155-156(1993).  
 CC -/- CATALYTIC ACTIVITY: D-GLUCOSE 6-PHOSPHATE + NADP(+) = D-GLUCONO-  
 CC DELTA-LACTONE 6-PHOSPHATE + NADPH.  
 CC -/- PATHWAY: FIRST STEP IN PENTOSE PHOSPHATE PATHWAY.  
 CC -/- SIMILARITY: BELONGS TO THE GLUCOSE-6-PHOSPHATE DEHYDROGENASE  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M55005; AAA24775.1; -;  
 DR EMBL; AE000279; AAC74922.1; -;  
 DR EMBL; D90828; BAA15660.1; -;  
 DR EMBL; U13783; AAA57018.1; -;  
 DR EMBL; U13784; AAA57019.1; -;  
 DR EMBL; U13785; AAA57020.1; -;  
 DR EMBL; U13786; AAA57021.1; -;  
 DR EMBL; U13787; AAA57022.1; -;  
 DR EMBL; U13788; AAA57023.1; -;  
 DR EMBL; U13789; AAA57024.1; -;  
 DR EMBL; U13790; AAA57025.1; -;  
 DR EMBL; U13791; AAA57026.1; -;  
 DR EMBL; U13792; AAA57027.1; -;  
 DR EMBL; U13793; AAA57028.1; -;  
 DR EMBL; U13794; AAA57029.1; -;  
 DR EMBL; X63694; CAA45220.1; -;  
 DR PIR; A38174; A38174.  
 DR HSSP; P11411; ZDPG.  
 DR SWISS-2DPAGE; F22992; COLI.  
 DR ECO2DBASE; F048.8; 6TH EDITION.  
 DR EcoGene; EGI1221; zwf.

DR InterPro; IPR001282; G6PD.  
 DR Pfam; PF00479; G6PD; 1.  
 DR PRINTS; PR00079; G6PDHGRGNASE.  
 DR PRODOM; PD001129; G6PD; 1.  
 DR PROSITE; PS00069; G6P\_DEHYDROGENASE; 1.  
 KW Oxidoreductase; NADP; Glucose metabolism; Complete proteome.  
 FT ACT\_SITE 181 BY SIMILARITY.  
 FT VARIANT 100 100 S -> N (IN STRAINS ECOR4 AND ECOR10).  
 FT CONFLICT 268 293 LKSLRRDRSNVREKTVRGVYTAGFA -> PEVSSPHRPLQ  
 FT SEQUENCE 491 AA; 55704 MW; 263F07D298EAFCD3 CRC64;  
 SQ  
 Query Match 36.1%; Score 43.5; DB 1; Length 491;  
 Best Local Similarity 69.2%; Pred. No. 25;  
 Matches 9; Conservative 1; Mismatches 2; Indels 1; Gaps 1;  
 QY 6 SIVEAWS-DNDTP 17  
 Db 451 SITEAWAMDNDAP 463  
 |||||  
 RESULT 15  
 YG33\_YEAST STANDARD; PRT; 275 AA.  
 ID YG33\_YEAST  
 AC P53271;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE HYPOTHETICAL 31.8 KDA PROTEIN IN NUP57-MEPI INTERGENIC REGION.  
 GN YGR120C OR G6324.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / FY1679;  
 RX MEDLINE=97197982; PubMed=9046098;  
 RA van Dyck L., Tettelin H., Purnelle B., Goffeau A.;  
 RT "An 18.3 kb DNA fragment from yeast chromosome VII carries four  
 RT unknown open reading frames, the gene for an Asn synthase, remnants  
 RT of Ty and three tRNA genes.";  
 RL Yeast 13:171-176(1997).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X83099; CAA58155.1; -;  
 DR EMBL; Z72905; CAA97130.1; -;  
 DR SGD; S0003352; SEC35.  
 KW Hypothetical protein.  
 SQ SEQUENCE 275 AA; 31799 MW; AA102D086FF3FAD7 CRC64;  
 Query Match 36.1%; Score 43; DB 1; Length 275;  
 Best Local Similarity 42.9%; Pred. No. 15;  
 Matches 6; Conservative 7; Mismatches 1; Indels 0; Gaps 0;  
 QY 7 IVEAWSNDTPYLH 20  
 Db 217 ILESCADSNSPYIH 230  
 |||||

Search completed: March 27, 2002, 14:27:05  
 Job time: 1647 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:26:06 ; Search time 188.53 Seconds  
(without alignments)  
16.293 Million cell updates/sec

Title: US-09-290-049a-14  
Perfect score: 119  
Sequence: 1 ANNHVSIVEAWSNDTPYLHD 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 14627329 residues  
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL\_17:\*
- 1: sp-archaea:\*
  - 2: sp-bacteria:\*
  - 3: sp-fungi:\*
  - 4: sp-human:\*
  - 5: sp-invertebrate:\*
  - 6: sp-mammal:\*
  - 7: sp-mhc:\*
  - 8: sp-organelle:\*
  - 9: sp-phage:\*
  - 10: sp-plant:\*
  - 11: sp-rodent:\*
  - 12: sp-virus:\*
  - 13: sp-vertebrate:\*
  - 14: sp-unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

alt No.	Score	Query Match	Length	DB ID	Description
1	119	100.0	1590	2 Q55263	Q55263 streptococc
2	119	100.0	1590	2 Q59983	Q59983 streptococc
3	108	90.8	1390	2 Q69385	Q69385 streptococc
4	108	90.8	1455	2 Q69391	Q69391 streptococc
5	99	83.2	1455	2 Q69382	Q69382 streptococc
6	99	83.2	1455	2 Q69388	Q69388 streptococc
7	99	83.2	1455	2 Q69397	Q69397 streptococc
8	76	63.9	1016	2 Q9LCJ7	Q9LCJ7 streptococc
9	68	57.1	1477	2 Q9LA66	Q9LA66 leuconostoc
10	68	57.1	1508	2 Q52224	Q52224 leuconostoc
11	68	57.1	1508	2 Q9EZH5	Q9EZH5 leuconostoc
12	68	57.1	1575	2 Q9LCH3	Q9LCH3 streptococc
13	68	57.1	1577	2 Q54178	Q54178 streptococc
14	67	56.3	1338	2 Q9WXJ4	Q9WXJ4 streptococc
15	65	54.6	1527	2 Q9ZAR4	Q9ZAR4 leuconostoc
16	61	51.3	1449	2 Q68542	Q68542 streptococc
17	61	51.3	1449	2 Q55264	Q55264 streptococc
18	60	50.4	1512	2 Q9WXJ5	Q9WXJ5 streptococc
19	57	47.9	1577	2 Q55265	Q55265 streptococc

20	56	47.1	1518	2 Q00600	Q00600 streptococc
21	55	46.2	1599	2 Q00599	Q00599 streptococc
22	55	46.2	2057	2 Q9RE05	Q9RE05 leuconostoc
23	53	44.5	175	10 Q9LMJ8	Q9LMJ8 arabidopsis
24	51	42.9	1195	2 Q06459	Q06459 thermus aqu
25	50	42.0	1252	5 Q9VQR4	Q9VQR4 drosophila
26	49	41.2	524	2 Q9KSI8	Q9KSI8 vibrio chol
27	47	39.5	193	2 Q9S445	Q9S445 pseudomonas
28	46.5	39.1	176	2 Q9KK16	Q9KK16 versinia en
29	46	38.7	271	11 Q9OZE9	Q9OZE9 rattus norv
30	46	38.7	490	2 Q06579	Q06579 mycobacteri
31	46	38.7	601	2 Q05566	Q05566 mycobacteri
32	46	38.7	880	5 P91643	P91643 drosophila
33	46	38.7	880	5 Q9VK54	Q9VK54 drosophila
34	45	37.8	172	10 Q9FJ22	Q9FJ22 arabidopsis
35	45	37.8	347	10 Q9LY92	Q9LY92 arabidopsis
36	45	37.8	525	3 Q59679	Q59679 schizosacch
37	45	37.8	947	10 Q23136	Q23136 arabidopsis
38	45	37.8	1366	4 Q9Y2G9	Q9Y2G9 homo sapien
39	45	37.8	1384	4 Q75257	Q75257 homo sapien
40	45	37.8	1683	5 Q9NKM1	Q9NKM1 leishmania
41	45	37.8	3895	12 Q65464	Q65464 border dise
42	45	37.8	3896	12 F87514	F87514 pestivirus
43	44.5	37.4	4848	2 Q07944	Q07944 streptomyc
44	44.5	37.4	835	2 Q25908	Q25908 helicobacte
45	44	37.0	136	12 Q9Q714	Q9Q714 classical s

ALIGNMENTS

RESULT 1	
Q55263	
ID Q55263	PRELIMINARY; PRT; 1590 AA.
AC Q55263;	
DT 01-NOV-1996 (TREMBlrel. 01, Created)	
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)	
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)	
DE GLUCOSYLTRANSFERASE GTF-I.	
OS Streptococcus sobrinus.	
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;	
CC Streptococcus.	
OX NCBI_TaxID=1310;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN=ATCC 33478;	
RA Sato S.;	
RL Ann. Kagoshima Univ. Dental School 16:23-29(1996).	
DR EMBL; D63570; BAA09792.1; -	
DR InterPro; IPR002479; CW_binding.	
DR InterPro; IPR003318; Glyco_hydro_70.	
DR Pfam; PF01473; CW_binding_I; 15.	
DR Pfam; PF02324; Glyco_hydro_70; 1.	
KW Transferase.	
SQ SEQUENCE 1590 AA; 176057 MW; 9DF7A3F2C6E4FD43 CRC64;	

Query Match 100.0%; Score 119; DB 2; Length 1590;  
Best Local Similarity 100.0%; Pred. No. 1.le-09;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1	ANNHVSIVEAWSNDTPYLHD 21
Db 477	ANNHVSIVEAWSNDTPYLHD 497

RESULT 2	
Q59983	
ID Q59983	PRELIMINARY; PRT; 1590 AA.
AC Q59983;	
DT 01-NOV-1996 (TREMBlrel. 01, Created)	
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)	

DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE  
DE 6-GLUCOSYLTRANSFERASE).

GN GTFI

OS Streptococcus sobrinus.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1310;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=OMZ176;

RX MEDLINE=94146405; PubMed=8312602;

RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;

RT "DNA sequence of the glucosyltransferase gene of serotype d

RT Streptococcus sobrinus.";

RL DNA Seq. 4:19-27(1993).

CC -I- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) = D-

CC EMBL; D13858; BAA02976.1; -

DR InterPro; IPR002479; CW\_binding.

DR Pfam; PF01473; CW\_binding\_1; 16

DR Pfam; PF02324; Glyco\_hydro\_70; 1.

KW Signal; Transferase; Glycosyltransferase.

FT SIGNAL 1 38 POTENTIAL.

FT CHAIN 39 1590 GLUCOSYLTRANSFERASE-I.

SQ SEQUENCE 1590 AA; 175955 MW; C3C83A57CF3C2B0E CRC64;

Query Match 100.0%; Score 119; DB 2; Length 1590;

Best Local Similarity 100.0%; Pred. No. 1.1e-09;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANNHVSIVEAWSNDNDTPYLHD 21

DB 477 ANNHVSIVEAWSNDNDTPYLHD 497

RESULT 3

ID O69385 PRELIMINARY; PRT; 1390 AA.

AC O69385;

DT 01-AUG-1998 (TrEMBLrel. 07, Created)

DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE GLUCOSYLTRANSFERASE-SI.

GN GTFC.

OS Streptococcus mutans.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1309;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MT4245;

RX MEDLINE=98231643; PubMed=9570124;

RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,

RA Kimura S., Hamada S.;

RT "Molecular analyses of glucosyltransferase genes among strains of

RT Streptococcus mutans.";

RL FEMS Microbiol. Lett. 161:331-336(1998).

DR EMBL; D88655; BAA26106.1; -

DR InterPro; IPR002479; CW\_binding.

DR InterPro; IPR003318; Glyco\_hydro\_70.

DR Pfam; PF01473; CW\_binding\_1; 7.

DR Pfam; PF02324; Glyco\_hydro\_70; 1.

KW Transferase.

SQ SEQUENCE 1390 AA; 155375 MW; 8847E4956EF05E9F CRC64;

Query Match 90.8%; Score 108; DB 2; Length 1390;

Best Local Similarity 85.7%; Pred. No. 4.7e-08;

Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANNHVSIVEAWSNDNDTPYLHD 21

Db 507 ANDHLSILEAWSNDNDTPYLHD 527

RESULT 4

ID O69391 PRELIMINARY; PRT; 1455 AA.

AC O69391;

DT 01-AUG-1998 (TrEMBLrel. 07, Created)

DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE GLUCOSYLTRANSFERASE-SI.

GN GTFC.

OS Streptococcus mutans.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1309;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MT4251;

RX MEDLINE=98231643; PubMed=9570124;

RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,

RA Kimura S., Hamada S.;

RT "Molecular analyses of glucosyltransferase genes among strains of

RT Streptococcus mutans.";

RL FEMS Microbiol. Lett. 161:331-336(1998).

DR EMBL; D88661; BAA26114.1; -

DR InterPro; IPR002479; CW\_binding.

DR InterPro; IPR003318; Glyco\_hydro\_70.

DR Pfam; PF01473; CW\_binding\_1; 9.

DR Pfam; PF02324; Glyco\_hydro\_70; 1.

KW Transferase.

SQ SEQUENCE 1455 AA; 162804 MW; 683A359D873E9E1A CRC64;

Query Match 90.8%; Score 108; DB 2; Length 1455;

Best Local Similarity 85.7%; Pred. No. 4.9e-08;

Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANNHVSIVEAWSNDNDTPYLHD 21

Db 507 ANDHLSILEAWSNDNDTPYLHD 527

RESULT 5

ID O69382 PRELIMINARY; PRT; 1455 AA.

AC O69382;

DT 01-AUG-1998 (TrEMBLrel. 07, Created)

DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE GLUCOSYLTRANSFERASE-SI.

GN GTFC.

OS Streptococcus mutans.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1309;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MT8148;

RX MEDLINE=98231643; PubMed=9570124;

RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,

RA Kimura S., Hamada S.;

RT "Molecular analyses of glucosyltransferase genes among strains of

RT Streptococcus mutans.";

RL FEMS Microbiol. Lett. 161:331-336(1998).

DR EMBL; D88652; BAA26102.1; -

DR InterPro; IPR002479; CW\_binding.

DR InterPro; IPR003318; Glyco\_hydro\_70.

DR Pfam; PF01473; CW\_binding\_1; 9.

DR Pfam; PF02324; Glyco\_hydro\_70; 1.

KW Transferase.

SQ SEQUENCE 1455 AA; 162969 MW; 27D4D3A1EECA2939 CRC64;

Query Match 83.2%; Score 99; DB 2; Length 1455;  
Best Local Similarity 81.0%; Pred. No. 1.2e-06;  
Matches 17; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ANNHVSIWEAWSNDTPYLHD 21  
|||:|||||  
Db 507 ANDHLSILEAWSNDTPYLHD 527

RESULT 6

ID O69388 PRELIMINARY; PRT; 1455 AA.  
AC O69388;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTFC.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OX Streptococcus.  
RN NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WT4239;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
RA Kimura S., Hamada S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of  
RT Streptococcus mutans";  
RL FEMS Microbiol. Lett. 161:331-336(1998).  
DR EMBL: D88658; BAA26110.1; -;  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam: PF01473; CW\_binding\_1; 10.  
DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1455 AA; 163045 MW; 6D90A4978D35DD82 CRC64;

Query Match 83.2%; Score 99; DB 2; Length 1455;  
Best Local Similarity 81.0%; Pred. No. 1.2e-06;  
Matches 17; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ANNHVSIWEAWSNDTPYLHD 21  
|||:|||||  
Db 507 ANDHLSILEAWSNDTPYLHD 527

RESULT 7

ID O69397 PRELIMINARY; PRT; 1455 AA.  
AC O69397;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTFC.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OX Streptococcus.  
RN NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WT4467;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
RA Kimura S., Hamada S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of  
RT Streptococcus mutans";  
RL FEMS Microbiol. Lett. 161:331-336(1998).

DR EMBL: D89978; BAA26120.1; -;  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam: PF01473; CW\_binding\_1; 9.  
DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1455 AA; 162913 MW; A1263427BF24E8E1 CRC64;

Query Match 83.2%; Score 99; DB 2; Length 1455;  
Best Local Similarity 81.0%; Pred. No. 1.2e-06;  
Matches 17; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ANNHVSIWEAWSNDTPYLHD 21  
|||:|||||  
Db 507 ANDHLSILEAWSNDTPYLHD 527

RESULT 8

Q9LCJ7  
ID Q9LCJ7 PRELIMINARY; PRT; 1016 AA.  
AC Q9LCJ7;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE DEXTRANSUCRASE.  
GN DSRT.  
OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OC Leuconostoc.  
OX NCBI\_TaxID=1245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NRRL B-512F;  
RX MEDLINE=20169623; PubMed=10705445;  
RA Funane K., Mizuno K., Takahara H., Kobayashi M.;  
RT "Gene encoding a dextranucrase-like protein in Leuconostoc  
RT mesenteroides NRRL B-512F";  
RL Biosci. Biotechnol. Biochem. 64:29-38(2000).  
DR EMBL: AB020020; BAA90527.1; -;  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
SQ SEQUENCE 1016 AA; 110343 MW; 8896EFDE13CCCB47 CRC64;

Query Match 63.9%; Score 76; DB 2; Length 1016;  
Best Local Similarity 61.9%; Pred. No. 0.0029;  
Matches 13; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 ANNHVSIWEAWSNDTPYLHD 21  
|||:|||||  
Db 548 SNQHVSILEDWSDNDAEYKVD 568

RESULT 9

Q9L466  
ID Q9L466 PRELIMINARY; PRT; 1477 AA.  
AC Q9L466;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE DEXTRANSUCRASE (EC 2.4.1.5).  
GN DSRC.  
OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OC Leuconostoc.  
OX NCBI\_TaxID=1245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NRRL B-1355;  
RA Arguello-Morales M.A., Renaud-Simeon M., Pizzut S., Sarcabal P.,  
RA Willemot R.M., Monsan P.;  
RT "Sequence analysis of the gene encoding alternansucrase, a sucrose

RT glucosyltransferase from *Leuconostoc mesenteroides* NRRL B-1355.";  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AJ250172; CAB76565.1; -;  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 14.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 DR KW Transferase; Glycosyltransferase.  
 SQ SEQUENCE 1477 AA; 164886 MW; E6F5710DEDFCB831 CRC64;

Query Match 57.1%; Score 68; DB 2; Length 1477;  
 Best Local Similarity 57.1%; Pred. No. 0.076;  
 Matches 12; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 ANNHVSIVEAWSNDTPYLHD 21  
 |||:||||| ||| ||| ||| |||

532 ANQHLSILEDWSHNDPEYVKD 552

RESULT 10

O52224 PRELIMINARY; PRT; 1508 AA.

ID O52224;  
 AC O52224;  
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE 6-  
 DE GLUCOSYLTRANSFERASE).  
 GN DSRB.

OS *Leuconostoc mesenteroides*.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC Leuconostoc.  
 OX NCBI\_TaxID=1245;

RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL B-1299;  
 RA Monchois V., Renaud-Simeon M., Monsan P., Willemot R.M.;  
 RL FEMS Microbiol. Lett. 0:0-0(1998).  
 CC -1- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) = D-  
 CC FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).  
 CC DR EMBL: AF301029; AAB95453.1; -;  
 CC InterPro: IPR002479; CW\_binding.  
 CC InterPro: IPR003318; Glyco\_hydro\_70.  
 CC Pfam: PF01473; CW\_binding\_1; 14.  
 CC Pfam: PF02324; Glyco\_hydro\_70; 1.  
 CC Transferase; Glycosyltransferase.

SEQUENCE 1508 AA; 168511 MW; E70CECB57A70D1F0 CRC64;  
 Query Match 57.1%; Score 68; DB 2; Length 1508;  
 Best Local Similarity 57.1%; Pred. No. 0.078;  
 Matches 12; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 ANNHVSIVEAWSNDTPYLHD 21  
 |||:||||| ||| ||| ||| |||

563 ANQHLSILEDWSHNDPEYVKD 583

RESULT 11

O9EZHS PRELIMINARY; PRT; 1508 AA.

ID O9EZHS;  
 AC O9EZHS;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE DEXTRANSUCRASE DSRB742.  
 DE DSRB742.  
 GN DSRB742.  
 OS *Leuconostoc mesenteroides*.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=B-742CB;  
 RA Kim H.-S., Kim D., Ryu H.-J., Robyt J.F.;  
 RT "Leuconostoc mesenteroides B-742CB, a dextranucrase gene.";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF294469; AAG38021.1; -;  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 14.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 SQ SEQUENCE 1508 AA; 168542 MW; E2FCFA0F87AE4F3A CRC64;

Query Match 57.1%; Score 68; DB 2; Length 1508;  
 Best Local Similarity 57.1%; Pred. No. 0.078;  
 Matches 12; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 ANNHVSIVEAWSNDTPYLHD 21  
 |||:||||| ||| ||| ||| |||

563 ANQHLSILEDWSHNDPEYVKD 583

RESULT 12

O9LCH3 PRELIMINARY; PRT; 1575 AA.

ID O9LCH3;  
 AC O9LCH3;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE.  
 GN GTFR.

OS *Streptococcus oralis*.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.

OX NCBI\_TaxID=1303;

RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC10557;  
 RA MEDLINE=20231779; PubMed=10768934;  
 RA Fujiwara T., Hoshino T., Sobue S., Hamada S.;  
 RT "Purification, characterization, and molecular analysis of the gene  
 RT encoding glucosyltransferase from *Streptococcus oralis*.";  
 RL Infect. Immun. 68:2475-2483(2000).  
 DR EMBL: AB025228; BAA95201.1; -;  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 17.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase.

SEQUENCE 1575 AA; 176792 MW; 772A26E4D7C2E543 CRC64;  
 Query Match 57.1%; Score 68; DB 2; Length 1575;  
 Best Local Similarity 61.9%; Pred. No. 0.082;  
 Matches 13; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 ANNHVSIVEAWSNDTPYLHD 21  
 |||:||||| ||| ||| ||| |||

546 AIKHLISLEAWSNDNDPDYVKD 566

RESULT 13

O54178 PRELIMINARY; PRT; 1577 AA.

ID O54178;  
 AC O54178; Q54247;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE.  
 GN STFG.  
 OS *Streptococcus gordonii* challis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.

OX NCBI\_TaxID=29390;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=CHALLIS;  
RX MEDLINE=96157084; PubMed=8586195;  
RA Vickerman M.M., Sulavik M.C., Clewell D.B.;  
RT "Molecular analysis of Streptococcus gordonii glucosyltransferase  
phase variants.";  
RL Dev. Biol. Stand. 85:309-314 (1995).  
RN [2]  
RP SEQUENCE OF 1-96 FROM N.A.  
RC STRAIN=CHALLIS;  
RX MEDLINE=92276337; PubMed=1534326;  
RA Sulavik M.C., Tardif G., Clewell D.B.;  
RT "Identification of a gene, *igt*, which regulates expression of  
glucosyltransferase and influences the Spp phenotype of Streptococcus  
gordonii Challis.";  
RL J. Bacteriol. 174:3577-3586 (1992).  
DR EMBL; U12643; AAC43483.1; -;  
DR EMBL; M89776; AAA26969.1; -;  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 18.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1577 AA; 177805 MW; 5AE0328DC5E08D18 CRC64;

Query Match 57.1%; Score 68; DB 2; Length 1577;  
Best Local Similarity 61.9%; Pred. No. 0.082;  
Matches 13; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 ANNHVSIVEAWSNDTPYLHD 21  
| : : : : : | : : |  
Db 548 ALKHLISILEAWSNDPDYKND 568

RESULT 14  
Q9WXJ4 PRELIMINARY; PRT; 1338 AA.  
AC Q9WXJ4;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE GTF-S.  
GN GTFs.  
OS Streptococcus criceti.  
OG Plasmid pAMI.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
NCBI\_TaxID=1333;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HS-6;  
RA Inoue M., Fukui K., Miyagi A.;  
RT "S. cricetus glucosyltransferase(*gtfs* and *gtft*) genes.";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB026123; BAA77236.1; -;  
DR HSP; P06278; IVDs.  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 10.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Plasmid.  
SQ SEQUENCE 1338 AA; 148558 MW; 0A90C8E10E15D99B CRC64;

Query Match 56.3%; Score 67; DB 2; Length 1338;  
Best Local Similarity 57.1%; Pred. No. 0.098;  
Matches 12; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 ANNHVSIVEAWSNDTPYLHD 21  
| : : : : : | : : |

Db 437 AIDHLSILEAWSNDNDYVKD 457  
RESULT 15  
Q9ZAR4 PRELIMINARY; PRT; 1527 AA.  
ID Q9ZAR4;  
AC Q9ZAR4;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE DEXTRANSUCRASE.  
GN DEX.  
OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OC Leuconostoc.  
OX NCBI\_TaxID=1245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NRRL B-512-F;  
RA Bhatnagar R., Singh D.K.S.;  
RT "Cloning and Molecular Characterization of Dextranucrase Gene from  
Leuconostoc mesenteroides NRRL B-512F.";  
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U81374; AAD10952.1; -;  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 16.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
SQ SEQUENCE 1527 AA; 169709 MW; 1DFAFA237C743398 CRC64;

Query Match 54.6%; Score 65; DB 2; Length 1527;  
Best Local Similarity 57.1%; Pred. No. 0.23;  
Matches 12; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 ANNHVSIVEAWSNDTPYLHD 21  
| : : : : : | : : |  
Db 581 ANQHLSILEAWSNDPLYVTD 601

Search completed: March 27, 2002, 14:26:07  
Job time: 1679 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:57:57 ; Search time 198.55 Seconds  
(without alignments)  
8.208 Million cell updates/sec

Title: US-09-290-049a-15  
Perfect score: 110  
Sequence: 1 VPSYFIRAHDSVDLIRNII 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

1 number of hits satisfying chosen parameters: 522463

Maximum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- A\_Geneseq\_ll01.\*
- 1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.\*
  - 2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.\*
  - 3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.\*
  - 4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.\*
  - 5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.\*
  - 6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.\*
  - 7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.\*
  - 8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.\*
  - 9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.\*
  - 10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.\*
  - 11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.\*
  - 12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.\*
  - 13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.\*
  - 14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.\*
  - 15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.\*
  - 16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.\*
  - 17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.\*
  - 18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.\*
  - 19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.\*
  - 20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.\*
  - 21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.\*
  - 22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	98	89.1	1592	AA32925	Glucosyltransferase
2	71	64.5	2057	AA10667	L. mesenteroides a
3	70	63.6	1577	AA91047	Alpha-D-glucosyltr
4	47	42.7	541	AA970433	Human atrial natri
5	47	42.7	541	AA95586	Atrial natriuretic
6	47	42.7	639	AA917089	Bacillus alkaline
7	45	40.9	1835	AA42839	Urea amidolysase.
8	44.5	40.5	986	AA937617	Protein which is s
9	44	40.0	344	AA40918	Bacillus subtilis
10	44	40.0	1068	AA43341	p110. AAR43341;
11	44	40.0	1068	AA43342	Human p110. Homo

12	44	40.0	1068	15	AA46294	PtdIns 3-kinase 11
13	43	39.1	164	21	AA44998	Zea mays protein f
14	43	39.1	190	21	AA44997	Zea mays protein f
15	43	39.1	210	21	AA44996	Zea mays protein f
16	43	39.1	347	21	AA86358	Human gene 11-enco
17	43	39.1	359	22	AA972339	Human protein #1 a
18	43	39.1	502	22	AA972340	Human protein #2 a
19	43	39.1	2541	21	AA841087	Human ORFX ORF851
20	42.5	38.6	451	22	AA41255	Human polypeptide
21	42	38.2	2625	19	AA55887	Human telomerase.
22	42	38.2	2627	19	AA61347	Human telomerase. R
23	41.5	37.7	713	17	AA999797	Lysine decarboxyla
24	41	37.3	403	21	AA42199	Arabidopsis thalia
25	41	37.3	510	21	AA42198	Arabidopsis thalia
26	41	37.3	518	21	AA42197	Arabidopsis thalia
27	41	37.3	519	20	AA48586	Human breast tumor
28	41	37.3	537	8	AA70432	Bovine atrial natri
29	41	37.3	537	17	AA95585	Atrial natriuretic
30	41	37.3	616	21	AA23861	Haemophilus influe
31	41	37.3	743	22	AA93069	Human protein sequ
32	41	37.3	743	22	AA93298	Human protein sequ
33	41	37.3	904	11	AA05619	BIV env and rev ex
34	41	37.3	1120	20	AA19780	B. burgdorferi ant
35	41	37.3	1123	21	AA58277	Heliothis armigera
36	41	37.3	1144	22	AA93293	Human polypeptide
37	41	37.3	1146	20	AA91979	B. burgdorferi ant
38	41	37.3	1165	22	AA939312	Human polypeptide
39	41	37.3	1177	22	AA41079	Human polypeptide
40	41	37.3	1177	22	AA41098	Human polypeptide
41	40.5	36.8	556	21	AA978522	Soybean ADA2 trans
42	40	36.4	43	20	AA925791	Human secreted pro
43	40	36.4	318	22	AAE01285	Human gene 12 enco
44	40	36.4	359	20	AA906263	RStA protein of CT
45	40	36.4	371	21	AA94076	Enzyme EPS2 involv

ALIGNMENTS

RESULT 1	
AA32925	AA32925 standard; Protein: 1592 AA.
ID	AA32925
XX	
AC	AA32925;
XX	
DT	28-JUN-1993 (first entry)
XX	
DE	Glucosyltransferase I.
XX	
KW	GT-1; Streptococcus; dental; caries.
XX	
OS	Streptococcus sobrinus.
XX	
FN	JP05023188-A.
XX	
PD	02-FEB-1993.
XX	
FF	25-JUL-1991; 91JP-0186592.
XX	
PR	25-JUL-1991; 91JP-0186592.
XX	
PA	(FUKU/) FUKUI I.
XX	
XX	(KATO/) KATO K.
DR	WPI; 1993-079449/10.
XX	
PT	N-PSDB; AAQ37760.
XX	
PT	DNA sequence glucosyl:transferase-I - comprises Streptococcus
XX	
PT	sobrinus DNA sequence with at least one nucleotide added or
XX	
PS	deleted
XX	
XX	Claim 13; Page 15; 29pp; Japanese.

CC The DNA sequence from Streptococcus sobrinus strain 6715 encodes  
 CC glucosyltransferase-1 (and mutants). The DNA was obtd. by treating  
 CC S. sobrinus 6715 with mutanolysin, extracting the chromosomal DNA,  
 CC partially digesting with Sau3AI and fractionating on agarose gel.  
 CC The 3-5 kbp fragment was ligated into pUC18 and E. coli JM109  
 CC transformed with it. A GT-1 expressing clone was isolated and  
 CC sequenced. The clone may be used in the development of a drug for  
 CC dental caries.  
 XX  
 SQ Sequence 1592 AA;

Query Match 89.1%; Score 98; DB 14; Length 1592;  
 Best Local Similarity 86.4%; Pred. No. 7.9e-08;  
 Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 VPSYSFTRAHDSEVDLIRNII 22  
 ||||| ||||| ||||| ||||| |||||  
 548 vpsysfarahdsevgdiirdii 569

RESULT 2  
 AAB10667  
 ID AAB10667 standard; Protein: 2057 AA.  
 XX  
 AC AAB10667;  
 XX  
 DT 19-JAN-2001 (first entry)  
 XX  
 DE L. mesenteroides alternan sucrose protein.  
 XX  
 KW Alternan sucrose; glucosyltransferase; fructose; cosmetic; foodstuff;  
 KW syrup.  
 XX  
 OS Leuconostoc mesenteroides.  
 XX  
 PN DE19905069-A1.  
 XX  
 PD 10-AUG-2000.  
 XX  
 PF 08-FEB-1999; 99DE-1005069.  
 XX  
 PR 08-FEB-1999; 99DE-1005069.  
 XX  
 PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.  
 PA (PLAC ) MAX PLANCK GES FORDERUNG WISSENSCHAFTEN.

Kossmann J, Welsh T, Quanz M, Knuth K;  
 WPI; 2000-550294/51.  
 N-PSDB; AAA97904.  
 XX  
 PT New nucleic acid encoding recombinant Leuconostoc mesenteroides  
 PT alternan sucrose protein and methods of alternan and fructose  
 PT production.  
 XX  
 PS Claim 1a; Page 30-36; 64pp; German.  
 XX  
 CC This invention describes a novel nucleic acid molecule (I) encoding an  
 CC alternan sucrose (E.C. 2.4.1.140 - an enzyme, that belongs to the  
 CC glucosyltransferase group) The recombinant, purified alternan sucrose  
 CC gene is useful for the fermentative production of alternan (a  
 CC carbohydrate) and/or fructose by secreting the enzyme into a  
 CC saccharose-containing culture medium. Alternatively, the enzyme is  
 CC contacted with a saccharose-containing solution. The alternan and/or  
 CC fructose is then isolated from the medium. Cosmetic products or  
 CC foodstuffs containing alternan can be produced. Recombinant production of  
 CC alternan sucrose is advantageous as it provides a cost effective means of  
 CC producing fructose for high fructose containing syrups, production of  
 CC which previously has been achieved by costly production from maize  
 CC starch. This sequence represents the Leuconostoc mesenteroides alternan  
 CC sucrose protein which is described in the method of the invention.  
 XX

SQ Sequence 2057 AA;  
 Query Match 64.5%; Score 71; DB 21; Length 2057;  
 Best Local Similarity 63.2%; Pred. No. 0.0044;  
 Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Oy 1 VPSYSFTRAHDSEVDLIR 19  
 :|||:||||: ||||  
 Db 757 ipnysfvrahdydaqdpir 775

RESULT 3  
 AAR91047  
 ID AAR91047 standard; Protein: 1577 AA.  
 XX  
 AC AAR91047;  
 XX  
 DT 22-MAY-1996 (first entry)  
 XX  
 DE Alpha-D-glucosyltransferase.  
 XX  
 KW Alpha-D-glucosyltransferase; primer-independent; soluble glucan;  
 KW sucrose; transgenic plant; cloning; Escherichia coli;  
 KW phage lambda-C13; vector; plasmid pGS501; plasmid pGS502;  
 KW gene transfer; crop improvement; storage carbohydrate; pasture;  
 KW feedstuff; senescence; dextran; binder; food; pharmaceutical.  
 XX  
 OS Streptococcus salivarius strain ATCC 25975.  
 XX  
 PN WO9606173-A1.  
 XX  
 PD 29-FEB-1996.  
 XX  
 PF 24-AUG-1995; 95WO-A000527.  
 XX  
 PR 24-AUG-1994; 94AU-0007643.  
 XX  
 PA (GIEFF) GIFFARD P M.  
 PA (JACQ) JACQUES N A.  
 PA (SIMP) SIMPSON C L.

Giffard PM, Jacques NA, Simpson CL;  
 WPI; 1996-151376/15.  
 N-PSDB; AAT13139.  
 XX  
 PT Plants contg. new bacterial DNA encoding glucosyl transferase  
 PT activity - retain higher levels of stored carbohydrate(s) in a form  
 PT readily digestible by ruminants  
 XX  
 PS Claim 4; Page 16-20; 31pp; English.  
 XX  
 CC The sequence represents an alpha-D-glucosyltransferase from  
 CC Streptococcus salivarius. The enzyme is primer-independent, and  
 CC produces soluble glucan from sucrose. A gene encoding the enzyme  
 CC may be cloned and expressed in Escherichia coli using a subclone  
 CC of phage lambda-C13, e.g. plasmid pGS501 or plasmid pGS502. The  
 CC DNA may also be expressed in a transgenic plant, to improve the  
 CC level of stored carbohydrate in a pasture plant which normally  
 CC contains low levels, or to prevent degradation of stored carbohydrate  
 CC during plant senescence. Dextran may be isolated from the plant, for  
 CC use as a food binder or pharmaceutical additive. Primer independence  
 CC ensures that the enzyme will be functional in plants. The glucan is  
 CC poorly degraded in plants but easily degraded by bacteria in the rumen  
 CC of grazing livestock.

SQ Sequence 1577 AA;  
 Query Match 63.6%; Score 70; DB 17; Length 1577;  
 Best Local Similarity 65.0%; Pred. No. 0.0048;  
 Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;



XX WPI; 1999-287736/24.  
DR N-PSDB; AAX37277.

XX  
XX Alkali protease from Bacillus used in washing powders

XX Disclosure; Page 53-58; 7lpp; Japanese.

XX The invention relates to alkaline proteases produced by strains of  
CC Bacillus. The proteases ability to digest casein is not inhibited by  
CC oleic acid and they have a high stability to oxidising agents. The  
CC alkaline protease of the invention has the following properties: (a) it  
CC is active over the pH range 4-13 and has at least 80% of its optimum  
CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is  
CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)  
CC its ability to digest casein is not inhibited by oleic acid; (e) it has  
CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be  
CC used as enzymes in washing compositions for use in automatic dishwashers  
CC and for washing clothes. The stability to oxidising agents allows the  
CC enzyme to be an effective component of washing compositions including  
CC bleaches. The present sequence represents an alkaline protease.

XX Sequence 639 AA;  
SQ

Query Match 42.7%; Score 47; DB 20; Length 639;  
Best Local Similarity 33.3%; Pred. No. 14;  
Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps

QY 1 VPSVSFIRAHDSVEVDLRNI 21  
DB 114 ipdayiveyegdvqskvrsi 134  
! ! : ! : ! : ! : ! : ! :

RESULT .7  
AAR42839  
ID AAR42839 standard; Protein; 1835 AA.  
XX AC AAR42839;  
XX AC AC  
XT 10-MAY-1994 (first entry)  
DT Urea amidolyase.  
DE Urea amidolyase:  
XX Urea amidolyase; URL; yeast; recombinant plasmid.  
XX Saccharomyces cerevisiae.  
PA JP05244959-A.  
PD 24-SEP-1993.  
XX PF 05-MAR-1992; 92JP-0084531.  
XX PR 05-MAR-1992; 92JP-0084531.  
XX PA (TOYM ) TOYOBO KK.  
XX WPI; 1993-338925/43.  
DR N-PSDB; AAQ49460.  
PT DNA having the genetic information of urea amidolyase originated  
TT from Saccharomyces yeast - can be used to prepare high purity  
XX urea amidolyase by culturing the transformant comprising the DNA  
XX Claim 2; Page 10-17; 17pp; Japanese.

XX This sequence represents a protein which has urea amidolyase (URL)-  
CC activity and is derived from yeast. The DNA encoding this protein  
CC may be used within a recombinant plasmid for the production of highly  
CC pure URL.  
XX Sequence 1835 AA;  
SQ

AA040918  
ID AAR40918 standard; Protein; 344 AA.

XX AC AAR40918;  
XX

DT 21-FEB-1994 (first entry)

XX DE Bacillus subtilis inositol dehydrogenase.

XX KW Recombinant protein production; inositol dehydrogenase.

XX OS Bacillus subtilis.

XX PN JP05192163-A.

XX PD 03-AUG-1993.

XX PF 30-SEP-1991; 91JP-0252073.

XX PR 30-SEP-1991; 91JP-0252073.

(SUNR ) SUNTORY LTD.

DR WPI; 1993-277476/35.

DR N-PSDB; AAQ48539.

XX PT Inositol dehydrogenase gene - isolated from e.g. bacteria  
PT belonging to Bacillus subtilis, in large amt.

XX PS Claim 2; Page 2; 18pp; Japanese.

XX CC The Bacillus subtilis inositol dehydrogenase gene has been cloned  
CC and sequenced. Microorganisms transformed by the coding sequence can  
CC be cultured to produce large amounts of recombinant inositol  
CC dehydrogenase.

XX SQ Sequence 344 AA;

Query Match 40.0%; Score 44; DB 14; Length 344;

Best Local Similarity 50.0%; Pred. No. 22;

Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 6 FIRAHSEVQDLIRNI 21

Db 276 fvaaydveiqdfidsi.291

RESULT 10

13341

AA043341 standard; Protein; 1068 AA.

AC AAR43341;

XX DT 12-APR-1994 (first entry)

XX DE p110.

XX KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;  
KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;  
KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides.

XX PN WO9321328-A.

XX PD 28-OCT-1993.

XX PF 13-APR-1993; 93WO-GB00761.

XX PR 13-APR-1992; 92GB-0008135.

XX PA (LUDW-) LUDWIG INST CANCER RES.

XX PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;

PI Parkerpj, Volinia S, Waterfield MD;

XX WPI; 1993-351738/44.

XX N-PSDB; AAQ51155.

XX PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase  
PT activity, useful for controlling cell proliferation

XX PS Claim 4; Fig 9; 146pp; English.

XX CC Southern blot analysis was performed using a bovine cDNA probe contg.  
CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated  
CC from a cDNA library constructed from mRNA isolated from the human  
CC cell line KGLa. Positive clones were sequenced to give a human  
CC PI3 kinase p110 sequence. This sequence has 95 percent  
CC homology with the bovine sequence. The domain encoding residues 19-  
CC 100 of human p110 is sufficient to encode the kinase which will  
CC associate with the p85 kinase subunit. The gene may be used to  
CC provide a protein with PI3 kinase activity, and is useful for  
CC screening for (ant)agonists of PI3 kinase activity which could be  
CC useful for stimulation or inhibition of cell proliferation and hence  
CC prophylaxis or therapy. Platelet or neutrophil activity or blood  
CC glucose levels can be controlled using the kinase.  
CC See also AAR43342 and AAR46552-3.

XX SQ Sequence 1068 AA;

Query Match 40.0%; Score 44; DB 14; Length 1068;

Best Local Similarity 45.5%; Pred. No. 84;

Matches 10; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

Qy 1 VPSYSFIRAHSEVQDLIRNII 22

Db 123 mpvcfdmvdvqdfvfrnii 144

RESULT 11

AAR43342

ID AAR43342 standard; Protein; 1068 AA.

XX AC AAR43342;

XX DT 12-APR-1994 (first entry)

XX DE Human p110.

XX KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;  
KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;  
KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT Domain 19..100

XX FT /note= "binds with p85 subunit"

XX PN WO9321328-A.

XX PD 28-OCT-1993.

XX PF 13-APR-1993; 93WO-GB00761.

XX PR 13-APR-1992; 92GB-0008135.

XX PA (LUDW-) LUDWIG INST CANCER RES.

XX PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;

XX PI Parkerpj, Volinia S, Waterfield MD;

XX WPI; 1993-351738/44.

XX N-PSDB; AAQ51155.

PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase  
 XX activity, useful for controlling cell proliferation  
 PS Claim 24; Fig 16; 146pp; English.

XX Southern blot analysis was performed using a bovine cDNA probe contg.  
 CC a fragment of a p13-kinase-encoding sequence and human cDNA isolated  
 CC from a cDNA library constructed from mRNA isolated from the human  
 CC cell line KG1a. Positive clones were sequenced to give the human  
 CC p13 kinase p110 sequence. This sequence has 95 percent homology  
 CC with the bovine sequence. The domain contg. residues 19-100 of human  
 CC p110 is sufficient to associate with the p85 kinase subunit. The  
 CC protein with p13 kinase activity is useful for screening for  
 CC (ant)agonists of p13 kinase activity which could be useful for  
 CC stimulation or inhibition of cell proliferation and hence  
 CC prophylaxis or therapy. Platelet or neutrophil activity or blood  
 CC glucose levels can be controlled using the kinase.  
 CC See also AAR43341 and AAR46552-3.

Sequence 1068 AA;

Query Match 40.0%; Score 44; DB 14; Length 1068;  
 Best Local Similarity 45.5%; Pred. No. 84;  
 Matches 10; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

Qy 1 VPSYSFIRAHDSVQDLIRNII 22  
 Db 123 mpvcfdmvpqvdfrnll 144

# RESULT 12

AAR46294  
 ID AAR46294 standard; Protein; 1068 AA.

XX AAR46294;

XX 31-AUG-1994 (first entry)

XX PtdIns 3-kinase 110 kD catalytic subunit.

XX 110 kD catalytic subunit; phosphatidyl inositol 3-kinase;  
 KW transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;  
 KW PtdIns 3-kinase; assay; detection; cell growth; regulation; cancer;  
 KW blood vessel plaques.

XX Bos taurus.

Key	Location/Qualifiers
FT Peptide	163..175
FT Peptide	/note= "Peptide A"
FT Peptide	326..337
FT Peptide	/note= "Peptide B"
FT Peptide	441..464
FT Peptide	/note= "Peptide C"
FT Peptide	622..648
FT Peptide	/note= "Peptide D"
FT Peptide	657..672
FT Peptide	/note= "Peptide E"
FT Peptide	679..699
FT Peptide	/note= "Peptide F"
FT Peptide	712..720
FT Peptide	/note= "Peptide G"
FT Peptide	868..882
FT Peptide	/note= "Peptide H"
FT Peptide	885..908
FT Peptide	/note= "Peptide I"
FT Peptide	925..941
FT Peptide	/note= "Peptide J"
FT Peptide	949..966
FT Peptide	/note= "Peptide K"
FT Peptide	987..1010
FT Peptide	/note= "Peptide L"

FT Peptide 1031..1040  
 FT Peptide /note= "Peptide M"  
 FT Peptide 1055..1063  
 FT Peptide /note= "Peptide N"

XX WO9403609-A.

XX 17-FEB-1994.

XX 05-AUG-1993; 93WO-GB01651.

XX 05-AUG-1992; 92GB-0016654.

XX (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.

XX Goode NT, Nurse PM, Parker PJJ, Waterfield MD;

XX WPI; 1994-065697/08.

XX N-PSDB; AAQ57012.

XX Eukaryotic cells transformed with mammalian phospholipid or  
 PT protein kinase DNA - useful in assays for compounds involved in  
 PT cell growth regulation and for treating cancers

PS Disclosure; Fig 1; 71pp; English.

XX This sequence represents the 110 kD catalytic subunit of the  
 CC phosphatidyl inositol (PtdIns) 3-kinase. The cDNA encoding this  
 CC sequence was transformed into Schizosaccharomyces pombe cells under  
 CC the regulatory control of the nmt promoter in an embodiment of the  
 CC invention. In the presence of thiamine the promoter is inactive and  
 CC the cells carrying the PtdIns catalytic subunit plasmid grow as the  
 CC parental strain. In the absence of thiamine the nmt promoter functions  
 CC and the PtdIns 3-kinase catalytic subunit is induced. PtdIns activity  
 CC is substantially increased under these conditions. Cells containing  
 CC constructs such as this, are useful in assays for detecting compounds  
 CC involved in cell growth regulation. It is also used as the basis for  
 CC detecting compounds for treating cancers and the formation of blood  
 CC vessel plaques.

XX Sequence 1068 AA;

Query Match 40.0%; Score 44; DB 15; Length 1068;  
 Best Local Similarity 45.5%; Pred. No. 84;  
 Matches 10; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

Qy 1 VPSYSFIRAHDSVQDLIRNII 22  
 Db 123 mpvcfdmvpqvdfrnll 144

# RESULT 13

AAG44998  
 ID AAG44998 standard; Protein; 164 AA.

XX AAG44998;

XX 18-OCT-2000 (first entry)

XX Zea mays protein fragment SEQ ID NO: 56436.

XX Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence; corn.

XX Zea mays subsp. mays.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136382.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.





```
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.

PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 39.1%; Score 43; DB 21; Length 190;
Best Local Similarity 47.1%; Pred. No. 16;
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 6 FIRAHSEVQDLIRNII 22
Db 96 fireytskvddlvkdxl 112
||| : | : | : | : | : |

RESULT 15
AAG44996
ID AAG44996 standard; Protein; 210 AA.
XX AC AAG44996;
XX DT 18-OCT-2000 (first entry)
XX DE Zea mays protein fragment SEQ ID NO: 56434.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX OS Zea mays subsp. mays.
XX PN EPL033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
```

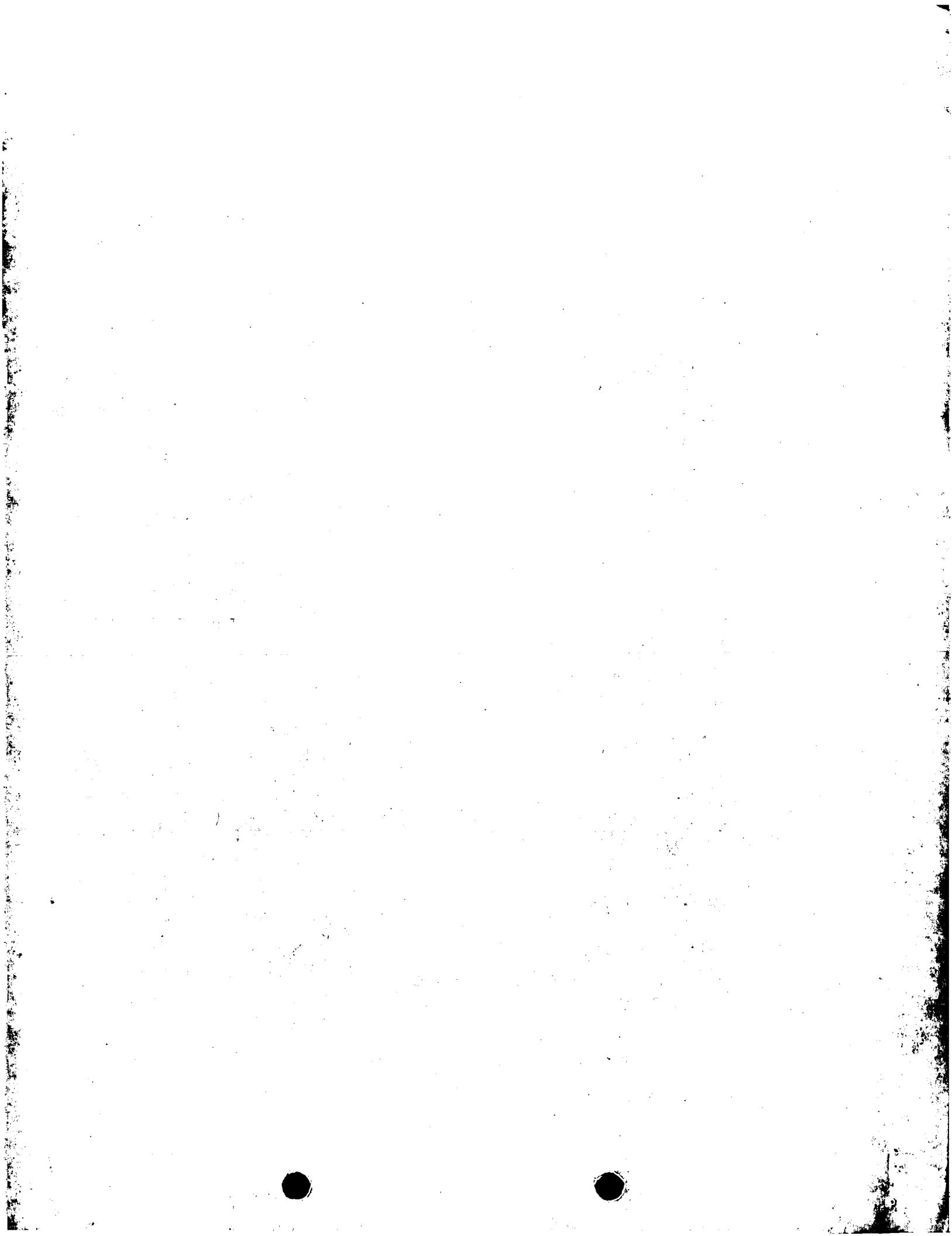
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134376.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142940.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0130891.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 18-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150366.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 31-AUG-1999; 99US-0151303.  
PR 01-SEP-1999; 99US-0151438.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 07-OCT-1999; 99US-0157865.  
PR 08-OCT-1999; 99US-0158029.  
PR 12-OCT-1999; 99US-0158232.  
PR 13-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.

PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match            39.1%; Score 43; DB 21; Length 210;  
Best Local Similarity    47.1%; Pred. No. 18;  
Matches    8; Conservative    5; Mismatches    4; Indels    0; Gaps    0;

QY         6 FIRAHDSVQDLIRNII 22  
||| : || : ||:: |
Db      116 fireytskvdvdkxi i32

Search completed: March 27, 2002, 13:57:58  
Job time: 526 sec





; CURRENT APPLICATION NUMBER: US/09/007,999  
; CURRENT FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-007-999-2

Query Match 90.0%; Score 99; DB 3; Length 1475;  
Best Local Similarity 90.9%; Pred. No. 1.8e-08;  
Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

1 VPSYSFIRAHSEVQDLIRNII 22  
|||||  
552 VPSYSFIRAHSEVQDLIADII 573

RESULT 3  
US-09-210-361-2  
; Sequence 2, Application US/09210361  
; Patent No. 6284479  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starches and  
; TITLE OF INVENTION: Latexes in Paper Manufacture  
; FILE REFERENCE: 0357CR  
; CURRENT APPLICATION NUMBER: US/09/210,361  
; CURRENT FILING DATE: 1998-12-11  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/009,620  
; EARLIER FILING DATE: 1998-01-20  
; EARLIER APPLICATION NUMBER: 08/485,243  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/008,172  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
EQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-210-361-2

Query Match 90.0%; Score 99; DB 4; Length 1475;  
Best Local Similarity 90.9%; Pred. No. 1.8e-08;  
Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSFIRAHSEVQDLIRNII 22  
|||||  
DB 552 VPSYSFIRAHSEVQDLIADII 573

RESULT 4  
US-08-793-824-2  
; Sequence 2, Application US/08793824  
; Patent No. 5981838  
; GENERAL INFORMATION:  
; APPLICANT: Simpson, Christine Lynn  
; APPLICANT: Gifford, Philip Morrison  
; APPLICANT: Jacques, Nicholas Anthony  
; TITLE OF INVENTION: Genetic Manipulation of Plants to  
; TITLE OF INVENTION: Increase Stored Carbohydrates

; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Griffith Hack & Co  
; STREET: Level 8, 168 Walker Street  
; CITY: No. 5981838th Sydney  
; STATE: New South Wales  
; COUNTRY: Australia  
; ZIP: 2060  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/793,824  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: AU PM7643  
; FILING DATE: 24-AUG-1994  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 61 2 9957 5944  
; TELEFAX: 61 2 957 6288  
; TELEX: 26547  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1577 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Streptococcus salivarius  
US-08-793-824-2

Query Match 63.6%; Score 70; DB 2; Length 1577;  
Best Local Similarity 65.0%; Pred. No. 0.0018;  
Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 SYSFIRAHSEVQDLIRNII 22  
:|:|||||:|  
DB 651 NVFVRAHSEVQAVLANII 680

RESULT 5  
US-09-008-172-2  
; Sequence 2, Application US/09008172  
; Patent No. 6127602  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; TITLE OF INVENTION: Latexes in Paper Manufacture  
; FILE REFERENCE: 0358D  
; CURRENT APPLICATION NUMBER: US/09/008,172  
; CURRENT FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-008-172-2

Query Match 61.8%; Score 68; DB 3; Length 1430;  
Best Local Similarity 70.0%; Pred. No. 0.0035;  
Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 SYSFIRAHSEVQDLIRNII 22  
:|:|||||:|

Db 576 NYIFRAHDSVQTVIAKII 595

## RESULT 6

; Sequence 6, Application US/09210361  
; Patent No. 6284479  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starches and  
; TITLE OF INVENTION: Latexes in Paper Manufacture  
; FILE REFERENCE: 0357CR  
; CURRENT APPLICATION NUMBER: US/09/210,361  
; CURRENT FILING DATE: 1998-12-11  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/009,620  
; EARLIER FILING DATE: 1998-01-20  
; EARLIER APPLICATION NUMBER: 08/485,243  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/008,172  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 6  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-210-361-6

Query Match 61.8%; Score 68; DB 4; Length 1430;  
Best Local Similarity 70.0%; Pred. No. 0.0035;  
Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 SYSFIRAHDSVQDLIRNII 22

Db 576 NYIFRAHDSVQTVIAKII 595

## RESULT 7

; Sequence 2, Application US/08684024  
; Patent No. 5834298  
; GENERAL INFORMATION:  
; APPLICANT: Benezra, Robert  
; TITLE OF INVENTION: GENE ENCODING THE HUMAN HOMOLOG OF MAD2  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION NUMBER: US/08/684,024  
; FILING DATE: 19-JUL-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 1747/46621-A  
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 278-0400  
; TELEFAX: (212) 391-0526  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 196 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-684-024-2

Query Match 41.8%; Score 46; DB 2; Length 196;  
Best Local Similarity 38.9%; Pred. No. 1.9;  
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 5 SFIRAHDSVQDLIRNII 22

Db 46 TLLKTHDDDELKDYIRKIL 63

## RESULT 8

US-08-684-024-9  
; Sequence 9, Application US/08684024  
; Patent No. 5834298  
; GENERAL INFORMATION:  
; APPLICANT: Benezra, Robert  
; TITLE OF INVENTION: GENE ENCODING THE HUMAN HOMOLOG OF MAD2  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION NUMBER: US/08/684,024  
; FILING DATE: 19-JUL-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 1747/46621-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 278-0400  
; TELEFAX: (212) 391-0526  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 196 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-684-024-9

Query Match 41.8%; Score 46; DB 2; Length 196;  
Best Local Similarity 38.9%; Pred. No. 1.9;  
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 5 SFIRAHDSVQDLIRNII 22

Db 46 TLLKTHDDDELKDYIRKIL 63

## RESULT 9

US-09-145-868-2

; Sequence 2, Application US/09145868  
; Patent No. 6096522  
; GENERAL INFORMATION:  
; APPLICANT: Benerra, Robert  
; TITLE OF INVENTION: GENE ENCODING THE HUMAN HOMOLOG OF MAD2  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/145.868  
; FILING DATE: 02-SEP-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 1747/46621-B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 278-0400  
; TELEFAX: (212) 391-0526  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 196 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-145-868-2

Query Match 41.8%; Score 46; DB 3; Length 196;  
Best Local Similarity 38.9%; Pred. No. 1.9;  
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 5 SFIRAHSEVDLIRNII 22  
Db 46 TLLKTHDDDELKDYIRKIL 63

RESULT 10  
; US-09-145-868-9  
; Sequence 9, Application US/09145868  
; Patent No. 6096522  
; GENERAL INFORMATION:  
; APPLICANT: Benerra, Robert  
; TITLE OF INVENTION: GENE ENCODING THE HUMAN HOMOLOG OF MAD2  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/145.868  
; FILING DATE: 02-SEP-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:

; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 1747/46621-B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 278-0400  
; TELEFAX: (212) 391-0526  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 196 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-145-868-9

Query Match 41.8%; Score 46; DB 3; Length 196;  
Best Local Similarity 38.9%; Pred. No. 1.9;  
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 5 SFIRAHSEVDLIRNII 22  
Db 46 TLLKTHDDDELKDYIRKIL 63

RESULT 11  
; US-08-390-874C-11  
; Sequence 11, Application US/08390874C  
; Patent No. 6043062  
; GENERAL INFORMATION:  
; APPLICANT: Klippel, Anke  
; ADDRESSEE: Williams, Lewis T.  
; TITLE OF INVENTION: A Constitutively Active  
; TITLE OF INVENTION: Phosphatidylinositol 3-Kinase and Uses Thereof  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/390.874C  
; FILING DATE: 17-FEB-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dow, Karen B.  
; REGISTRATION NUMBER: 29,684  
; REFERENCE/DOCKET NUMBER: 02307K-0570000US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1068 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-390-874C-11

Query Match 40.0%; Score 44; DB 3; Length 1068;  
Best Local Similarity 45.5%; Pred. No. 33;  
Matches 10; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

Qy 1 VPSYSFIRAHSEVDLIRNII 22











```
Qy      1 VPSYFIRAHDSVQDLIRNII 22  
        |||||  
Db      1 VPSYFIRAHDSVQDLIRNII 22
```

RESULT 2  
US-09-740-274-4  
; Sequence 4, Application US/09740274  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 1375  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-740-274-4

Query Match 100.0%; Score 110; DB 21; Length 1375;  
Best Local Similarity 100.0%; Pred. No. 7.8e-09;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPSYSFIRAHSEVQDLIRNII 22  
|||||  
DB 578 VPSYSFIRAHSEVQDLIRNII 599

RESULT 3  
US-09-290-049-17  
; Sequence 17, Application US/09290049  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: FDC98-0102A  
; CURRENT APPLICATION NUMBER: US/09/290,049  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; EARLIER FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; EARLIER FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 17  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. downei  
US-09-290-049-17

Query Match 90.9%; Score 100; DB 16; Length 22;  
Best Local Similarity 90.9%; Pred. No. 2e-09;  
Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSFIRAHSEVQDLIRNII 22  
|||||  
DB 1 VPSYSFIRAHSEVQDLIRNII 22

RESULT 4  
US-09-557-848-2  
; Sequence 2, Application US/09557848  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; FILE REFERENCE: 0356D2  
; CURRENT APPLICATION NUMBER: US/09/557,848  
; CURRENT FILING DATE: 2000-04-26  
; PRIOR APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-557-848-2

Query Match 90.0%; Score 99; DB 19; Length 1475;  
Best Local Similarity 90.9%; Pred. No. 6.3e-07;  
Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSFIRAHSEVQDLIRNII 22  
|||||  
DB 552 VPSYSFIRAHSEVQDLIRNII 573

RESULT 5  
US-09-740-274-2  
; Sequence 2, Application US/09740274  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-740-274-2

Query Match 90.0%; Score 99; DB 21; Length 1475;  
Best Local Similarity 90.9%; Pred. No. 6.3e-07;  
Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSFIRAHSEVQDLIRNII 22  
|||||  
DB 552 VPSYSFIRAHSEVQDLIRNII 573

```

RESULT 6
US-09-290-049-19
; Sequence 19, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: FDC98-01p2A
; CURRENT APPLICATION NUMBER: US/09/290,049
; EARLIER FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 19
; LENGTH: 22
; TYPE: PRT
; ORGANISM: S. sobrinus
US-09-290-049-19

```

```

Query Match      89.1%; Score 98; DB 16; Length 22;
Best Local Similarity 86.4%; Pred. No. 4.4e-09;
Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 VPSYSFIRAHDSVEVDLIRNII 22
||||| ||||| ||||| ||||| |||||
Db 1 VPSYSFIRAHDSVEVDLIRNII 22

```

```

RESULT 7
US-09-290-049-2
; Sequence 2, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: FDC98-01p2A
; CURRENT APPLICATION NUMBER: US/09/290,049
; EARLIER FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: HDS peptide
US-09-290-049-2

```

```

Query Match      73.2%; Score 80.5; DB 16; Length 20;
Best Local Similarity 94.7%; Pred. No. 3.6e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```

```

Qy 1 VPSYSFIR-AHDSVEVDLI 18
||||| ||||| ||||| ||||| |||||
Db 1 VPSYSFIRAHDSVEVDLI 19

```

```

RESULT 8
US-09-290-049-18
; Sequence 18, Application US/09290049

```

```

; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: FDC98-01p2A
; CURRENT APPLICATION NUMBER: US/09/290,049
; EARLIER FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 22
; TYPE: PRT
; ORGANISM: S. downei
US-09-290-049-18

```

```

Query Match      69.1%; Score 76; DB 16; Length 22;
Best Local Similarity 72.7%; Pred. No. 2.4e-05;
Matches 16; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```

```

Qy 1 VPSYSFIRAHDSVEVDLIRNII 22
||||| ||||| ||||| ||||| |||||
Db 1 VPSYSFIRAHDSVEVDLIRNII 22

```

```

RESULT 9
US-09-499-203-2
; Sequence 2, Application US/09499203
; GENERAL INFORMATION:
; APPLICANT: KOSMANN, Jens
; APPLICANT: WELSH, Thomas
; APPLICANT: QUANZ, Martin
; APPLICANT: KNUTH, Karola
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase
; FILE REFERENCE: 147-196P
; CURRENT APPLICATION NUMBER: US/09/499,203
; CURRENT FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2057
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-499-203-2

```

```

Query Match      64.5%; Score 71; DB 18; Length 2057;
Best Local Similarity 63.2%; Pred. No. 0.054;
Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy 1 VPSYSFIRAHDSVEVDLIR 19
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 757 IPNYSFVRADHYDAQDPIR 775

```

```

RESULT 10
US-09-290-049-16
; Sequence 16, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: FDC98-01p2A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13

```

; EARLIER APPLICATION NUMBER: 60/115,142  
; EARLIER FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 16  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. mutans  
US-09-290-049-16

Query Match 61.8%; Score 68; DB 16; Length 22;  
Best Local Similarity 70.0%; Pred. No. 0.00054;  
Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 3 SYSFIRAHDSVQDLIRNII 22  
:|||||:|:|  
3 NYIFIRAHDSVQTVIAKII 22

RESULT 11  
US-09-649-885-2  
; Sequence 2, Application US/09649885  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; FILE REFERENCE: 035822  
; CURRENT APPLICATION NUMBER: US/09/649,885  
; CURRENT FILING DATE: 2000-08-28  
; PRIOR APPLICATION NUMBER: US 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-649-885-2

Query Match 61.8%; Score 68; DB 20; Length 1430;  
Best Local Similarity 70.0%; Pred. No. 0.11;  
Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 3 SYSFIRAHDSVQDLIRNII 22  
:|||||:|:|  
Db 576 NYIFIRAHDSVQTVIAKII 595

RESULT 12  
US-09-740-274-6  
; Sequence 6, Application US/09740274  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16

; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 6  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-740-274-6

Query Match 61.8%; Score 68; DB 21; Length 1430;  
Best Local Similarity 70.0%; Pred. No. 0.11;  
Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 3 SYSFIRAHDSVQDLIRNII 22  
:|||||:|:|  
Db 576 NYIFIRAHDSVQTVIAKII 595

RESULT 13  
US-09-417-507-43716  
; Sequence 43716, Application US/09417507  
; GENERAL INFORMATION:  
; APPLICANT: KEITH G. WEINSTOCK ET AL.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ASPERGILLUS  
; FILE REFERENCE: PATH99-10  
; CURRENT APPLICATION NUMBER: US/09/417,507  
; CURRENT FILING DATE: 1999-10-14  
; NUMBER OF SEQ ID NOS: 44312  
; SEQ ID NO 43716  
; LENGTH: 194  
; TYPE: PRT  
; ORGANISM: A. fumigatus  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (5),(8),(56)  
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are un  
US-09-417-507-43716

Query Match 45.5%; Score 50; DB 18; Length 194;  
Best Local Similarity 47.4%; Pred. No. 9.7;  
Matches 9; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 2 PSYSFIRAHDSVQDLIRN 20  
||:|:|:|:|:|:|:|  
Db 14 PSFSFRSRSRKEVREAMRN 32

RESULT 14  
US-09-107-532-4217  
; Sequence 4217, Application US/09107532  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7308  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD/ROM ISO9660  
; COMPUTER:  
; OPERATING SYSTEM:  
; SOFTWARE:  
; CURRENT APPLICATION DATA:



; APPLICATION NUMBER: US/09/107,532  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/ 085598  
; FILING DATE: May 14, 1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 4217:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 565 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: 1...565  
; US-09-107-532-4217

Query Match 45.5%; Score 50; DB 15; Length 565;  
Best Local Similarity 36.8%; Pred. No. 38;  
Matches 7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 2 PSYSFIRAHDSVDLIRN 20  
||:|: || |: :|||  
Db 480 PSFAVATHDIELTEILKN 498

## RESULT 15

US-09-107-532A-4217  
; Sequence 4217, Application US/09107532A  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007

; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 4217:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 565 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (B) LOCATION 1...565  
; SEQUENCE DESCRIPTION: SEQ ID NO: 4217:  
US-09-107-532A-4217

Query Match 45.5%; Score 50; DB 15; Length 565;  
Best Local Similarity 36.8%; Pred. No. 38;  
Matches 7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 2 PSYSFIRAHDSVDLIRN 20  
||:|: || |: :|||  
Db 480 PSFAVATHDIELTEILKN 498

Search completed: March 27, 2002, 14:20:28  
Job time: 1575 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:22:48 ; Search time 137.48 Seconds  
(without alignments)  
11.042 Million cell updates/sec

Title: US-09-290-049A-15  
Perfect score: 110  
Sequence: 1 VPSYSFIRAHSEVDLIRNII 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 332938 seqs, 68999538 residues

Total number of hits satisfying chosen parameters: 332938

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_New.\*

- 1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep1.\*
- 7: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep.\*
- 8: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	110	100.0	22	6	US-09-562-328-28
2	110	100.0	22	6	US-09-290-049A-15
3	100	90.9	22	6	US-09-562-328-30
4	100	90.9	22	6	US-09-290-049A-17
5	98	89.1	22	6	US-09-562-328-32
6	98	89.1	22	6	US-09-290-049A-19
7	91	82.7	19	6	US-09-562-328-27
8	91	82.7	19	6	US-09-290-049A-2
9	91	82.7	23	6	US-09-562-328-44
10	82	74.5	523	6	US-09-604-957-5
11	80.5	73.2	20	6	US-09-562-328-26
12	76	69.1	22	6	US-09-562-328-31
13	76	69.1	22	6	US-09-290-049A-18
14	76	69.1	535	6	US-09-604-957-7
15	76	69.1	1278	6	US-09-604-957-3
16	68	61.8	22	6	US-09-562-328-29
17	68	61.8	22	6	US-09-290-049A-16
18	68	61.8	545	6	US-09-604-957-4
19	66	60.0	584	6	US-09-604-957-6
20	55	50.0	215	6	US-09-675-784A-8997
21	47	42.7	639	6	US-09-920-954-4
22	46.5	42.3	389	6	US-09-708-427-10452
23	46.5	42.3	398	6	US-09-708-427-10451
24	46.5	42.3	414	6	US-09-708-427-10450
25	44	40.0	327	8	US-60-356-051-2881

26	43	39.1	274	6	US-09-675-784A-9123	Sequence 9123, Ap
27	43	39.1	347	7	US-10-012-542-277	Sequence 277, App
28	43	39.1	347	7	US-10-050-873-281	Sequence 281, App
29	43	39.1	555	6	US-09-708-427-7685	Sequence 7685, Ap
30	43	39.1	591	6	US-09-708-427-7684	Sequence 7684, Ap
31	43	39.1	635	6	US-09-708-427-7683	Sequence 7683, Ap
32	43	39.1	832	6	US-09-708-427-3256	Sequence 3256, Ap
33	42.5	38.6	444	6	US-09-614-150-20343	Sequence 20343, A
34	42	38.2	95	6	US-09-897-516-8009	Sequence 8009, Ap
35	42	38.2	2535	6	US-09-708-427-26651	Sequence 26651, A
36	41.5	37.7	1213	1	PCT-US02-03987-5358	Sequence 5358, Ap
37	41.5	37.7	1213	6	US-09-815-242-5358	Sequence 5358, Ap
38	41.5	37.7	1213	7	US-10-072-851-5358	Sequence 5358, Ap
39	41.5	37.7	1217	1	PCT-US02-03987-12523	Sequence 12523, A
40	41.5	37.7	1217	6	US-09-815-242-12523	Sequence 12523, A
41	41.5	37.7	1217	7	US-10-072-851-12523	Sequence 12523, A
42	41	37.3	188	6	US-09-675-784A-10453	Sequence 10453, A
43	41	37.3	425	6	US-09-708-427-11088	Sequence 11088, A
44	41	37.3	441	6	US-09-708-427-11087	Sequence 11087, A
45	41	37.3	521	6	US-09-646-673A-117	Sequence 117, App

ALIGNMENTS

RESULT 1

US-09-562-328-28  
; Sequence 28, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995-0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 28  
; LENGTH: 22  
; TYPE: PNT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-28

Query Match 100.0%; Score 110; DB 6; Length 22;  
Best Local Similarity 100.0%; Pred. No. 3.9e-11;  
Matches 22 Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPSYS.FIRAHSEVDLIRNII 22  
Db 1 VPSYS.FIRAHSEVDLIRNII 22

RESULT 2

US-09-290-049A-15  
; Sequence 15, Application US/09290049A  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CARIES  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 15  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. mutans  
US-09-290-049a-15

Query Match 100.0%; Score 110; DB 6; Length 22;  
Best Local Similarity 100.0%; Pred. No. 3.9e-11;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPSYSFIRAHDSVQDLIRNII 22  
Db 1 VPSYSFIRAHDSVQDLIRNII 22

RESULT 3  
US-09-562-328-30  
Sequence 30, Application US/09562328  
GENERAL INFORMATION:

; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 30  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-30

Query Match 90.9%; Score 100; DB 6; Length 22;  
Best Local Similarity 90.9%; Pred. No. 1.5e-09;  
Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VPSYSFIRAHDSVQDLIRNII 22  
Db 1 VPSYSFIRAHDSVQDLIRNII 22

RESULT 4  
US-09-290-049a-17  
Sequence 17, Application US/09290049a  
GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CARRIES  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049a  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 17  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. downei  
US-09-290-049a-17

Query Match 90.9%; Score 100; DB 6; Length 22;  
Best Local Similarity 90.9%; Pred. No. 1.5e-09;

Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 VPSYSFIRAHDSVQDLIRNII 22  
Db 1 VPSYSFIRAHDSVQDLIRNII 22

RESULT 5  
US-09-562-328-32  
Sequence 32, Application US/09562328  
GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 32  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-32

Query Match 89.1%; Score 98; DB 6; Length 22;  
Best Local Similarity 86.4%; Pred. No. 3.2e-09;  
Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VPSYSFIRAHDSVQDLIRNII 22  
Db 1 VPSYSFIRAHDSVQDLIRNII 22

RESULT 6  
US-09-290-049a-19  
Sequence 19, Application US/09290049a  
GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CARRIES  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049a  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 19  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: S. sobrinus  
US-09-290-049a-19

Query Match 89.1%; Score 98; DB 6; Length 22;  
Best Local Similarity 86.4%; Pred. No. 3.2e-09;  
Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VPSYSFIRAHDSVQDLIRNII 22  
Db 1 VPSYSFIRAHDSVQDLIRNII 22

RESULT 7  
US-09-562-328-27

; Sequence 27, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 27  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-27

Query Match 82.7%; Score 91; DB 6; Length 19;  
Best Local Similarity 100.0%; Pred. No. 3.5e-08; Indels 0; Gaps 0;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPSYSFIRAHDSVDLI 18  
| | | | | | | | | | | | | | | | | |  
Db 1 VPSYSFIRAHDSVDLI 18

RESULT 8  
US-09-290-049A-2

; Sequence 2, Application US/09290049A  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: HDS peptide  
US-09-290-049A-2

Query Match 82.7%; Score 91; DB 6; Length 19;  
Best Local Similarity 100.0%; Pred. No. 3.5e-08; Indels 0; Gaps 0;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPSYSFIRAHDSVDLI 18  
| | | | | | | | | | | | | | | | | |  
Db 1 VPSYSFIRAHDSVDLI 18

RESULT 9  
US-09-562-328-44

; Sequence 44, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01

; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 44  
; LENGTH: 23  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-44

Query Match 82.7%; Score 91; DB 6; Length 23;  
Best Local Similarity 100.0%; Pred. No. 4.4e-08; Indels 0; Gaps 0;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPSYSFIRAHDSVDLI 18  
| | | | | | | | | | | | | | | | | |  
Db 5 VPSYSFIRAHDSVDLI 22

RESULT 10  
US-09-604-957-5

; Sequence 5, Application US/09604957  
; GENERAL INFORMATION:  
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITINA HENDRIKA  
; APPLICANT: DIJKHUIZEN, LOBBERT  
; APPLICANT: RAHAOUT, HAKIM  
; APPLICANT: LEER, ROBERT-JAN  
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
; FILE REFERENCE: BO 43388  
; CURRENT APPLICATION NUMBER: US/09/604,957  
; CURRENT FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 00201871.1  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 523  
; TYPE: PRT  
; ORGANISM: Leuconostoc mesenteroides  
US-09-604-957-5

Query Match 74.5%; Score 82; DB 6; Length 523;  
Best Local Similarity 63.6%; Pred. No. 5.7e-05; Indels 3; Gaps 0;  
Matches 14; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VPSYSFIRAHDSVDLIRNII 22  
| | | | | | | | | | | | | | | | | |  
Db 146 IPNYSFIRAHDSVDLIRNII 167

RESULT 11  
US-09-562-328-26

; Sequence 26, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 26  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.

US-09-562-328-26

Query Match	73.28;	Score 80.5;	DB 6;	Length 20;
Best Local Similarity	94.78;	Pred. No. 1.8e-06;		

```
QY      1 VPSYSFIR-AHDSEVQDLI 18
        | | | | | | | | | | | |
Db      1 VPSYSFIRTAHDSEVQDLI 19
```

RESULT 12  
US-09-562-328-31  
; Sequence 31, Application US/09562328

; APPLICANT: LEES, ANDREW  
 ; APPLICANT: TAUBMAN, MARTIN A.  
 ; APPLICANT: SMITH, DANIEL J.  
 ; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
 ; FILE REFERENCE: 04995.0046-01  
 ; CURRENT APPLICATION NUMBER: US/09/562,328  
 ; CURRENT FILING DATE: 2000-05-01  
 ; PRIOR APPLICATION NUMBER: 09/288,965  
 ; PRIOR FILING DATE: 1999-04-09  
 ; NUMBER OF SEQ ID NOS: 44  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 31  
 ; LENGTH: 22  
 ; TYPE: PR1  
 ; ORGANISM: Streptococcus sp.  
 ; US-09-562-328-31

Query Match 69.1%; Score 76; DB 6; Length 22;  
Best Local Similarity 72.7%; Pred. NO. 1e-05;  
Matches 16; Conservative 1; Mismatches 5; Indels

QY 1 VPSYSFIRAHDSVQDLIRNII 22  
||:| ||||| || ||

Db 1 VPNYVFIRAHDSVQTRIAKII 22

RESULT 13  
US-09-290-049A-18  
Sequence 18, Application US/092900049A

```

1  APPLICANT: Smith, Daniel J.
2  APPLICANT: Fauman, Martin A.
3  TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
4  TITLE OF INVENTION: CARRIES
5  FILE REFERENCE: 1564.1008-002
6  CURRENT APPLICATION NUMBER: US/09/290,049A
7  CURRENT FILING DATE: 1999-04-12
8  PRIOR APPLICATION NUMBER: 60/081,550
9  PRIOR FILING DATE: 1998-04-13
10 PRIOR APPLICATION NUMBER: 60/115,142
11 PRIOR FILING DATE: 1999-01-08
12 NUMBER OF SEQ ID NOS: 19
13 SOFTWARE: FastSeq for Windows Version 4.0
14 SEQ ID NO 18
15 LENGTH: 22
16 TYPE: PRT
17 ORGANISM: S. downei
18 US-09-290-049A-18

```

Query Match 69.1%; Score 76; DB 6; Length 22;  
Best Local Similarity 72.7%; Pred. No. 1e-05;  
Matches 16; Conservative 1; Mismatches 5; Indels

Qy 1 VPSYFIRAHDSVQDLIRNII 22  
||:| ||||| ||| | ||

1 VPYVFIRAH DSEVOTRIAKII 22

```

RESULT 14
US-09-604-957-7
; Sequence 7, Application US/09604957
; GENERAL INFORMATION:
; APPLICANT: VAN GEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 4338
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
; US-09-604-957-7

```

Query Match	69.1%	Score 76;	DB 6;	Length 535;
Best Local Similarity	59.1%	Pred. No. 0.00053;		
Matches 13;	Conservative	5;	Mismatches 4;	Indels 0;
Gaps				

```
Qy 1 VPSYFIRAHDSVQDLIRNII 22
    :|:|:|:|:|:|:|:|:|:|:|
Db 144 IPNYSFVRAHDNNSQDQIQNAI 165
```

```

RESULT 15
US-09-604-957-3
; Sequence 3, Application US/09604957
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 4388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1278
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-604-957-3

```

Query Match	69.1%	Score 76;	DB 6;	Length 1278;
Best Local Similarity	59.1%;	Pred. No. 0.0015;		
Matches 13;	Conservative	5;	Mismatches 4;	Indels 0;
Matches 13;	Conservative	5;	Mismatches 4;	Indels 0;

```
Qy 1 VPSYFIRAHDSVQDLIRNII 22
      :|:|:|:|:|:|:|:|:|:|:|
Db 620 IPNYSEVRAHDNNSODOIONAI 641
```

Search completed: March 27, 2002, 14:22:48  
Job time: 1695 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:01:21 ; Search time 102.51 Seconds  
(without alignments)  
16.348 Million cell updates/sec

Title: US-09-290-049a-15

Perfect score: 110

Sequence: 1 VPSYSFIRAHDSVDLIRNII 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_68.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	110	100.0	1375	2 JT0345	dextranucrase (EC
2	99	90.0	1475	2 B33135	gtfB protein precu
3	98	89.1	1592	2 A38175	glucosyltransferas
4	83	75.5	1290	2 JC5473	dextranucrase (EC
5	83	75.5	1508	2 T31098	probable dextranu
6	76	69.1	1365	2 A41483	glucosyltransferas
7	71	64.5	1449	2 T30857	glucosyltransferas
8	71	64.5	1449	2 T30552	glucosyltransferas
9	70	63.6	1577	2 T30858	glucosyltransferas
10	68	61.8	1431	2 A45866	dextranucrase (EC
11	68	61.8	1518	2 A44811	glucosyltransferas
12	68	61.8	1599	2 S22737	glucosyltransferas
13	50	45.5	51	2 G82455	hypothetical prote
14	48	43.6	540	1 OYHUCR	natriuretic peptid
15	46.5	42.3	414	2 T06120	hypothetical prote
16	46	41.8	196	2 S48302	MAD2 protein - yea
17	46	41.8	597	2 F82935	DNA polymerase III
18	45.5	41.4	767	2 T39715	probable transcrip
19	45.5	41.4	956	2 H81654	conserved hypotet
20	45	40.9	135	2 T12823	hypothetical prote
21	45	40.9	1835	2 S46082	urea carboxylase (
22	45	40.9	6359	2 T31679	bacitracin synthet
23	44.5	40.5	954	2 G71496	hypothetical prote
24	44	40.0	202	2 E72688	hypothetical prote
25	44	40.0	282	2 S47795	probable transcrip
26	44	40.0	327	2 H59094	hypothetical prote
27	44	40.0	327	2 D86474	probable RING zinc
28	44	40.0	344	2 JH0511	myo-inositol 2-deh
29	44	40.0	587	2 S63033	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

JT0345

dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)  
N:Alternate names: sucrose 6-glucosyltransferase  
C:Species: Streptococcus mutans  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 24-Sep-1999  
C:Accession: JT0345; C33135

R:Ueda, S.; Shiroza, T.; Kuramitsu, H. K.

Gene 69, 101-109, 1988

A:Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.

A:Reference number: JT0345; MUID:89137980

A:Accession: JT0345

A:Molecule type: DNA

A:Residues: 1-1375 <UED>

A:Experimental source: GS-5

R:Shiroza, T.; Ueda, S.; Kuramitsu, H. K.

J. Bacteriol. 169, 4263-4270, 1987

A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.

A:Reference number: A33135; MUID:87308013

A:Accession: C33135

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-349 <SHI>

A:Cross-references: GB:M17361

C:Genetics:

A:Gene: gtfC

C:Function:

A:Description: catalyzes the synthesis of both water-soluble and water-insoluble gluc

C:Superfamily: cpl repeat homology

C:Keywords: dextranation; glucosyltransferase; hexosyltransferase

F:1-34/Domain: total sequence #status predicted <SIG>

F:33-1375/Product: glucosyltransferase #status predicted <MAT>

F:1126-1145/Domain: cpl repeat homology <CP1>

F:1253-1272/Domain: cpl repeat homology <CP2>

F:1318-1337/Domain: cpl repeat homology <CP3>

Query Match 100.0%; Score 110; DB 2; Length 1375;

Best Local Similarity 100.0%; Pred. No. 3.9e-09;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPSYSFIRAHDSVDLIRNII 22

|||||

Db 578 VPSYSFIRAHDSVDLIRNII 599

|||||

##### RESULT 2

B33135

gtfB protein precursor - Streptococcus mutans

C:Species: Streptococcus mutans

C:Date: 23-Oct-1990 #sequence\_revision 23-Oct-1990 #text\_change 15-Oct-1999

C:Accession: B33135; A33128



## RESULT 9

A44811  
 glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius  
 C:Species: Streptococcus salivarius  
 C:Date: 31-Dec-1993 #sequence: revision 31-Dec-1993 #text\_chan  
 C:Accession: A44811; S22726; S28809  
 R:Giffard, P. M.; Simpson, C.L.; Milward, C. P.; Jacques, N.A.  
 J. Gen. Microbiol. 137, 2577-2593, 1991  
 A:Title: Molecular characterization of a cluster of at least  
 A:Reference number: A44811; MUID:92148377

A:Accession: A44811  
A:Molecule type: DNA  
A:Residues: 1-1518  
A:Cross-references: NCBI  
A:Note: sequence extracted from  
C:Genetics:  
A:Gene: gtfJ  
C:Superfamily: cpl re  
C:Keywords: glycosyl  
F:1307-1326/Domain: c

Query Match 61.8%; Score 68; DB 2; Length 1518;  
Best Local Similarity 65.0%; Pred. NO. 0.023;  
Matches 13; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

3 SYSFIRAHDSVQDLIRNII 22
: | | | | | | | | | | | |
604 NYVFIRAHDNNVODIIAEII 623

```

RESULT 12  
S22737  
glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius  
C:Species: Streptococcus salivarius  
C:Date: 16-Sep-1992 #sequence revision 16-Sep-1992 #text change 21-Jan-2000

Query Match 61.8%; Score 68; DB 2; Length 1599;  
Best Local Similarity 65.0%; Pred. NO. 0.024;  
Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```
Qy      3 SYSFIRAHDSVQDLIRNII 22
        : | | | | | | | | | | : | |
Db     574 TYLFVRAHDSEVQTVIADII 593
```

RESULT 13  
G82455  
hypothetical protein VCA0471 [imported] - Vibrio cholerae (strain N16961 serogroup O1)  
C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C:Accession: G82455  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
Richardson, D.; Esmolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.  
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833  
A:Accession: G82455  
A:Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-51 <HEL>  
A:Cross-references: GB:AE004379; GB:AE003853; NID:9957865; PIDN:AAF96375.1; GSPDB:GN  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VCA0471  
A:Map position: 2

```

Query Match      45.5%; S
Best Local Similarity 54.5%; P
Matches 12: Conservative 5;

Qy      1 VPSYSFIRAHDSVQDLRNII 22
Db      15 VPVSAIR--NSEIRLKEII 34

```

RESULT 14  
OYHUCR  
natriuretic peptide receptor C precursor - human  
N:Alternate names: atrial natriuretic factor clearance receptor; natriuretic peptide  
C:Species: Homo sapiens (man)  
C:Date: 09-Nov-1990 #sequence\_revision 05-May-1995 #text\_change 22-Jun-1999  
C:Accession: S10150; A35896  
R:Lowy, D.G.; Camerato, T.R.; Goeddel, D.V.  
Nucleic Acids Res. 18, 3412, 1990  
A:Title: cDNA sequence of the human atrial natriuretic peptide clearance receptor.  
A:Reference number: S10150; MUID:90287735

RESULT 15  
T06120  
hypothetical protein F23E12.70 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 22-Oct-1999  
C:Accession: T06120  
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Hobeisel,  
submitted to the Protein Sequence Database, April 1999  
A:Reference number: Z15485  
A:Accession: T06120  
A:Molecule type: DNA  
A:Residues: 1-414 <BEV>  
A:Cross-references: EMBL:AL022604; GSPDB:GN00062; ATSP:F23E12.70  
A:Experimental source: cultivar Columbia; BAC clone F23E12  
C:Genetics:  
A:Gene: ATSP:F23E12.70  
A:Map position: 4  
A:Introns: 83/1; 109/3; 180/1; 199/3; 215/3; 259/3; 284/3; 317/3; 359/3

Query Match 42.3%; Score 46.5; DB 2; Length 414;  
Best Local Similarity 55.0%; Pred. No. 14;  
Matches 11; Conservative 2; Mismatches 6; Indels 1; Gaps 1;

QY 2 PSYSFIRAHDSVQDLIRNI 21  
||: | ||||| : ||  
Db 337 PSF-IIHAHDSVSSISYNI 355

Search completed: March 27, 2002, 14:01:22  
Job time: 484 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:59:31 ; Search time 87.3 seconds  
(without alignments)  
5.413 Million cell updates/sec

Title: US-09-290-049A-14  
Perfect score: 119  
Sequence: 1 ANNHVSIVEAWSNDTPYLHD 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

- 1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/iaa/PCTUS.COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	108	90.8	1475	3	US-09-007-999-2
2	108	90.8	1475	4	US-09-210-361-2
3	99	83.2	1375	4	US-09-210-361-4
4	74	62.2	1430	3	US-09-008-172-2
5	74	62.2	1430	4	US-09-210-361-6
6	57	47.9	1577	2	US-08-793-824-2
7	42	35.3	200	2	US-08-836-442-4
8	42	35.3	334	6	5290690-11
9	42	35.3	455	4	US-09-362-473-6
10	42	35.3	529	2	US-08-836-442-3
11	42	35.3	1627	1	US-07-665-792E-9
12	42	35.3	3898	4	US-08-750-717-2
13	41	34.5	259	1	US-08-277-231A-3
14	41	34.5	259	2	US-08-473-750-6
15	41	34.5	259	2	US-08-477-326-6
16	41	34.5	2465	3	US-08-596-291-3
17	41	34.5	2465	3	US-09-100-804-3
18	41	34.5	2466	3	US-09-080-855-12
19	41	34.5	2466	5	PCT-US94-09943-2
20	41	34.5	2485	4	US-09-290-640-46
21	41	34.5	3898	2	US-08-876-991-2
22	41	34.5	3898	2	US-09-059-853-2
23	40.5	34.0	214	4	US-09-214-278-1
24	40.5	34.0	1055	4	US-09-214-278-2
25	40.5	34.0	1148	4	US-08-882-046-4
26	40.5	34.0	1212	4	US-09-214-278-3
27	40.5	34.0	1238	4	US-09-214-278-5

28	40.5	34.0	1248	4	US-08-882-046-6	Sequence 6, Appli
29	40	33.6	196	2	US-08-933-750C-35	Sequence 35, Appl
30	40	33.6	196	4	US-09-234-613-35	Sequence 35, Appl
31	40	33.6	700	3	US-08-931-952-2	Sequence 2, Appli
32	40	33.6	700	3	US-08-272-247-2	Sequence 2, Appli
33	40	33.6	700	5	PCT-US95-08560-2	Sequence 2, Appli
34	40	33.6	839	4	US-09-197-636-2	Sequence 2, Appli
35	40	33.6	839	4	US-09-197-636-4	Sequence 2, Appli
36	40	33.6	839	4	US-09-197-636-8	Sequence 8, Appli
37	39	32.8	82	4	US-09-073-297-22	Sequence 22, Appli
38	39	32.8	384	3	US-09-311-170-2	Sequence 2, Appli
39	39	32.8	385	1	US-08-361-920-23	Sequence 23, Appl
40	39	32.8	385	1	US-08-479-939-23	Sequence 23, Appl
41	39	32.8	385	1	US-08-483-432-23	Sequence 23, Appl
42	39	32.8	445	5	PCT-US94-05387-7	Sequence 7, Appli
43	39	32.8	895	1	US-08-123-161A-8	Sequence 8, Appli
44	39	32.8	895	1	US-08-483-278-8	Sequence 8, Appli
45	39	32.8	1091	3	US-08-633-768A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-09-007-999-2  
; Sequence 2, Application US/09007999  
; Patent No. 6087559  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; FILE REFERENCE: 0356D  
; CURRENT APPLICATION NUMBER: US/09/007,999  
; CURRENT FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-007-999-2

Query Match 90.8%; Score 108; DB 3; Length 1475;  
Best Local Similarity 85.7%; Pred. No. 8.2e-09;  
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANNHVSIVEAWSNDTPYLHD 21  
Db 481 ANDHLSILEAWSNDTPYLHD 501

RESULT 2  
US-09-210-361-2  
; Sequence 2, Application US/09210361  
; Patent No. 6284479  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starches and  
; FILE REFERENCE: 0357CE  
; CURRENT APPLICATION NUMBER: US/09/210,361  
; CURRENT FILING DATE: 1998-12-11  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/009,620  
; EARLIER FILING DATE: 1998-01-20  
; EARLIER APPLICATION NUMBER: 08/485,243  
; EARLIER FILING DATE: 1995-06-07







APPLICANT: Kinney, Anthony J.  
TITLE OF INVENTION: Aromatic Amino Acid Catabolism Enzymes  
FILE REFERENCE: BB-1197  
CURRENT APPLICATION NUMBER: US/09/362,473  
CURRENT FILING DATE: 1999-07-28  
EARLIER APPLICATION NUMBER: 60/094,783  
EARLIER FILING DATE: JULY 31, 1998  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 6  
LENGTH: 455  
TYPE: PRT  
ORGANISM: Glycine max  
US-09-362-473-6

Query Match 35.3%; Score 42; DB 4; Length 455;  
Best Local Similarity 50.0%; Pred. No. 57;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 8 VEASNDTPLYLHD 21  
DB 270 VVAHGNVPMYD 283

RESULT 10  
US-08-836-442-3  
Sequence 3, Application US/08836442  
Patent No. 5990293  
GENERAL INFORMATION:  
APPLICANT: DOCHERTY, Andrew, J. P.  
APPLICANT: SLOCOMBE, Patrick, M.  
TITLE OF INVENTION: A HUMAN METALLOPROTEINASE  
TITLE OF INVENTION: VARIANTS THEREOF AND DNA SEQUENCES CODING THEREFOR  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP  
STREET: 130 Water Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/836,442  
FILING DATE: 01-MAY-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB96/02101  
FILING DATE: 13-MAR-1997  
APPLICATION NUMBER: GB 9612150.4  
FILING DATE: 11-JUN-1996  
APPLICATION NUMBER: GB 9526229.1  
FILING DATE: 21-DEC-1995  
APPLICATION NUMBER: GB 9521498.7  
FILING DATE: 20-OCT-1995  
APPLICATION NUMBER: GB 95521495.3  
FILING DATE: 20-OCT-1995  
APPLICATION NUMBER: GB 9518023.8  
FILING DATE: 05-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Resnick, David S  
REGISTRATION NUMBER: 34,235  
REFERENCE/DOCKET NUMBER: 47425  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-523-3400  
TELEFAX: 617-523-6440  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:  
LENGTH: 529 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: No. 5990293e  
US-08-836-442-3

Query Match 35.3%; Score 42; DB 2; Length 529;  
Best Local Similarity 50.0%; Pred. No. 68;  
Matches 8; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 2 NNHVSVIV--EAWSDND 15  
DB 246 NTHVALVGMEIWTDKD 261

RESULT 11  
US-07-665-792E-9  
Sequence 9, Application US/07665792E  
Patent No. 5281694  
GENERAL INFORMATION:  
APPLICANT: Baseman, Joel B.  
APPLICANT: Su, C. J.  
APPLICANT: Dallo, S. F.  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR  
TITLE OF INVENTION: PRODUCTION OF MYOPLASMAL ADHESINS  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARNOLD, WHITE & DURKEE  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: U.S.A.  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5-1/4" DSDD Diskette  
COMPUTER: Compaq (IBM PC Compatible)  
OPERATING SYSTEM: MS DOS 3.31  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/665,792E  
FILING DATE: 19910607  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/118,967  
FILING DATE: No. 5281694. 10, 1987  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/588,886  
FILING DATE: July 27, 1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/616,111  
FILING DATE: No. 5281694. 21, 1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/697,349  
FILING DATE: May 8, 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Melinda Patterson  
REGISTRATION NUMBER: 33,062  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (713) 787-1592  
TELEFAX: (713) 789-2679  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1627 Amino Acids  
TYPE: AMINO ACID  
TOPOLOGY: Linear  
MOLECULE TYPE: Genomic DNA  
DESCRIPTION: Deduced amino acid sequence of p1  
DESCRIPTION: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:



NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/473,750  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/277,321  
FILING DATE: 19-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Alice O.  
REGISTRATION NUMBER: 33,542  
REFERENCE/DOCKET NUMBER: ACC94-02B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 259 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-473-750-6

Query Match 34.5%; Score 41; DB 2; Length 259;  
Best Local Similarity 37.5%; Pred. No. 43;  
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 2 NNHVSIVEAWSNDNTP 17  
:: :|:| | | |  
Db 72 DDSAALVQAWIDNGNP 87

## RESULT 15

US-08-477-326-6  
Sequence 6, Application US/08477326  
Patent No. 5968769  
GENERAL INFORMATION:  
APPLICANT: Green, Bruce A.  
APPLICANT: Brinton, Jr., Charles C.  
TITLE OF INVENTION: Sequence and Analysis of LKP Filin  
Patent No. 5968769  
TITLE OF INVENTION: Structural Gene and the LKP Pili Operon of No. 5968769typable  
TITLE OF INVENTION: Haemophilus Influenzae  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,326  
FILING DATE: 07-JUN-1995  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/277,231  
FILING DATE: July 19, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Alice O.  
REGISTRATION NUMBER: 33,542  
REFERENCE/DOCKET NUMBER: ACC94-02A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 259 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-477-326-6

Query Match 34.5%; Score 41; DB 2; Length 259;  
Best Local Similarity 37.5%; Pred. No. 43;  
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 2 NNHVSIVEAWSNDNTP 17  
:: :|:| | | |  
Db 72 DDSAALVQAWIDNGNP 87

Search completed: March 27, 2002, 13:59:31  
Job time: 584 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:20:27 ; Search time 1139.61 Seconds  
(without alignments)  
5.116 Million cell updates/sec

Title: US-09-290-049A-14  
Perfect score: 119  
Sequence: 1 ANNHVSIVEAWSNDTPYLHD 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 3148936 seqs, 277657034 residues

Total number of hits satisfying chosen parameters: 3148936

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: /cgn2\_6/ptodata/2/paa/PCTUS\_COMB.pep.\*
  - 2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep.\*
  - 3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep.\*
  - 4: /cgn2\_6/ptodata/2/paa/US08\_COMB.pep.\*
  - 5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pep.\*
  - 6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pep.\*
  - 7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pep.\*
  - 8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep.\*
  - 9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pep.\*
  - 10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep.\*
  - 11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pep.\*
  - 12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pep.\*
  - 13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep.\*
  - 14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep.\*
  - 15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pep.\*
  - 16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pep.\*
  - 17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pep.\*
  - 18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pep.\*
  - 19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pep.\*
  - 20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pep.\*
  - 21: /cgn2\_6/ptodata/2/paa/US097\_COMB.pep.\*
  - 22: /cgn2\_6/ptodata/2/paa/US098\_COMB.pep.\*
  - 23: /cgn2\_6/ptodata/2/paa/US099\_COMB.pep.\*
  - 24: /cgn2\_6/ptodata/2/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	119	100.0	21	16	US-09-290-049-12
2	119	100.0	21	16	US-09-290-049-14
3	108	90.8	21	16	US-09-290-049-1
4	108	90.8	21	16	US-09-290-049-10
5	108	90.8	1475	19	US-09-557-848-2
6	108	90.8	1475	21	US-09-740-274-2
7	99	83.2	1375	21	US-09-740-274-4
8	74	62.2	21	16	US-09-290-049-11
9	74	62.2	1430	20	US-09-649-885-2
					Sequence 12, Appl
					Sequence 14, Appl
					Sequence 1, Appl
					Sequence 10, Appl
					Sequence 2, Appl
					Sequence 2, Appl
					Sequence 4, Appl
					Sequence 11, Appl
					Sequence 2, Appl

10	74	62.2	1430	21	US-09-740-274-6	Sequence 6, Appl
11	62	52.1	21	16	US-09-290-049-13	Sequence 13, Appl
12	55	46.2	2057	18	US-09-499-203-2	Sequence 2, Appl
13	53	44.5	196	19	US-09-595-298A-765	Sequence 765, App
14	50	42.0	1252	24	US-60-191-637-4899	Sequence 4899, Ap
15	48.5	40.8	401	1	PCT-US01-01309-222	Sequence 222, App
16	47	39.5	749	19	US-09-562-737-96	Sequence 96, Appl
17	46.5	39.1	220	24	US-60-212-413-164	Sequence 164, App
18	46.5	39.1	220	24	US-60-229-518-331	Sequence 331, App
19	46.5	39.1	330	24	US-60-207-422-75	Sequence 75, Appl
20	46.5	39.1	539	1	PCT-US01-01332-762	Sequence 762, App
21	46.5	39.1	539	21	US-09-764-875-762	Sequence 704, App
22	46	38.7	47	21	US-09-758-470-704	Sequence 31546, A
23	46	38.7	164	18	US-09-417-507-31546	Sequence 2, Appl
24	46	38.7	880	15	US-09-137-970-2	Sequence 12250, A
25	46	38.7	880	24	US-60-167-217-12250	Sequence 9936, Ap
26	46	38.7	880	24	US-60-173-464-9936	Sequence 12251, A
27	46	38.7	880	24	US-60-191-637-12251	Sequence 9608, Ap
28	46	38.7	880	24	US-60-191-681-9608	Sequence 4571, Ap
29	45	37.8	199	21	US-09-738-626-4571	Sequence 661, App
30	45	37.8	306	19	US-09-595-298A-661	Sequence 660, App
31	45	37.8	340	19	US-09-595-298A-660	Sequence 1370, Ap
32	45	37.8	1270	24	US-60-242-679-1370	Sequence 22573, A
33	44	37.0	73	16	US-09-248-796-22573	Sequence 21844, A
34	44	37.0	129	16	US-09-248-796-21844	Sequence 13119, A
35	44	37.0	261	23	US-09-902-540-13119	Sequence 20736, A
36	44	37.0	391	24	US-60-324-109-20736	Sequence 4985, Ap
37	44	37.0	659	24	US-60-167-217-4985	Sequence 11, Appl
38	43.5	36.6	202	10	US-08-633-719-11	Sequence 11, Appl
39	43.5	36.6	202	16	US-09-280-197-11	Sequence 9810, Ap
40	43.5	36.6	656	23	US-09-902-540-9810	Sequence 3, Appl
41	43.5	36.6	1092	12	US-08-836-156-3	Sequence 3, Appl
42	43.5	36.6	1092	16	US-09-375-608-3	Sequence 3, Appl
43	43.5	36.6	1092	18	US-09-423-126-5	Sequence 891, App
44	43	36.1	40	24	US-60-196-190-891	Sequence 324, App
45	43	36.1	67	20	US-09-684-524-324	

ALIGNMENTS

RESULT 1  
US-09-290-049-12  
; Sequence 12, Application US/09290049  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: PDC98-01b2A  
; CURRENT APPLICATION NUMBER: US/09/290,049  
; EARLIER FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081,550  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/115,142  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 12  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: S. downei  
US-09-290-049-12

Query Match	100.0%	Score 119;	DB 16;	Length 21;
Best Local Similarity	100.0%	Pred. No. 9.6e-12;		
Matches	21;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0;
Qy	1	ANNHVSIVEAWSNDTPYLHD 21		
Db	1	ANNHVSIVEAWSNDTPYLHD 21		





```

; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2

```

```

Query Match          90.8%; Score 108; DB 21; Length 1475;
Best Local Similarity 85.7%; Pred. No. 7.4e-08;
Matches 19; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ANNHVSIVEAWSNDTPYLHD 21
   |::|::|::|::|::|::|::|
Db 481 ANDHLSILEAWSNDTPYLHD 501

```

```

RESULT 7
US-09-740-274-4
; Sequence 4, Application US/09740274
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-4

```

```

Query Match          83.2%; Score 99; DB 21; Length 1375;
Best Local Similarity 81.0%; Pred. No. 2e-06;
Matches 17; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 ANNHVSIVEAWSNDTPYLHD 21
   |::|::|::|::|::|::|::|
Db 507 ANDHLSILEAWSNDTPYLHD 527

```

```

RESULT 8
US-09-290-049-11
; Sequence 11, Application US/09290049

```

```

; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CARRIES
; FILE REFERENCE: FDC98-01p2A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. mutans
US-09-290-049-11

```

```

Query Match          62.2%; Score 74; DB 16; Length 21;
Best Local Similarity 66.7%; Pred. No. 0.00018;
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 ANNHVSIVEAWSNDTPYLHD 21
   |::|::|::|::|::|::|
Db 1 AINHLSILEAWSNDTPQYNKD 21

```

```

RESULT 9
US-09-649-885-2
; Sequence 2, Application US/09649885
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0358D2
; CURRENT APPLICATION NUMBER: US/09/649,885
; CURRENT FILING DATE: 2000-08-28
; PRIOR APPLICATION NUMBER: US 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 143;
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-649-885-2

```

```

Query Match          62.2%; Score 74; DB 20; Length 1430;
Best Local Similarity 66.7%; Pred. No. 0.023;
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 ANNHVSIVEAWSNDTPYLHD 21
   |::|::|::|::|::|::|
Db 495 AINHLSILEAWSNDTPQYNKD 515

```

```

RESULT 10
US-09-740-274-6
; Sequence 6, Application US/09740274
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11

```

```

; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
; 9-740-274-6

```

Query Match 62.2%; Score 74; DB 21; Length 1430;  
Best Local Similarity 66.7%; Pred. NO. 0.023;  
Matches 14; Conservative 2; Mismatches 5; Indels

Qy 1 ANNHVSIVEAWSNDNTPYLHD 21  
| | | : | | | | |  
Db 495 AINHLSILEAWSNDNPYKND 515

```

RESULT 11
US-09-290-049-13
; Sequence 13, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACC
; TITLE OF INVENTION: CARRIES
; FILE REFERENCE: FDC98-01p2A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. downei
US-09-290-049-13

```

Query Match 52.1%; Score 62; DB 16; Length 21;  
Best Local Similarity 57.9%; Pred. No. 0.016;  
Matches 11; Conservative 4; Mismatches 4; Indels

**Qy**      1 ANNHVSIVEAWSNDTPYL 19  
         | :|:|:|:| | | :

**Dd**      1 AIDHLSILEAWSGNDNDYV 19

RESULT 12  
US-09-499-203-2  
; Sequence 2, Application US/09499203  
; GENERAL INFORMATION:  
; APPLICANT: KOSSMANN, Jens  
; APPLICANT: WELSH, Thomas  
; APPLICANT: QUANZ, Martin  
; APPLICANT: KNUTH, Karola  
; TITLE OF INVENTION: Nucleic Acid M  
; FILE REFERENCE: 147-196P

```

; CURRENT APPLICATION NUMBER: US/09/499, 203
; CURRENT FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 2
; LENGTH: 2057
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-499-203-2

```

Query Match 46.2%; Score 55; DB 18; Length 2057;  
Best Local Similarity 45.0%; Pred. No. 40;  
Matches 9; Conservative 5; Mismatches 6; Indels

```
Qy 1 ANNHVSIVEAWSNDNTPYLH 20
    |||::|||:| |::
Db 665 ANKHLSELDWNGKDPQVYN 684
```

```

RESULT 13
US-09-595-298A-765
; Sequence 765, Application US/09595298A
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai
; APPLICANT: BROVER, Vyacheslav
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypept
; TITLE OF INVENTION: Thereby
; FILE REFERENCE: 2750-0953p
; CURRENT APPLICATION NUMBER: US/09/595, 298A
; CURRENT FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 2756
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 765
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Arabidopsis Thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: (1)..(196)
; OTHER INFORMATION: Ceres Seq. ID no. 1026995
; NAME/KEY: misc_feature
; LOCATION: ()..()
; OTHER INFORMATION: Xaa is any aa, unknown or other
US-09-595-298A-765

```

Query Match 44.5%; Score 53; DB 19; Length 196;  
Best Local Similarity 42.1%; Pred. No. 5.8;  
Matches 8; Conservative 4; Mismatches 7; Indels

```

QY      3 NHVSIVEAWSNDNTPYLHD 21
      || ::|||: | |
Db     109 NHQEVIDAWSDHQKPLWTD 127

```

```

RESULT 14
US-60-191-637-4899
; Sequence 4899, Application US/60191637
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING
; TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000392
; CURRENT APPLICATION NUMBER: US/60/191.637
; CURRENT FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 42660
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4899
; LENGTH: 1252
; TYPE: PRT
; ORGANISM: DROSOPHILA

```

US-60-191-637-4899

Query Match 42.0%; Score 50; DB 24; Length 1252;  
Best Local Similarity 38.1%; Pred. No. 1.5e+02;  
Matches 8; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

Qy 1 ANNHVSIVEAWSNDNDTPYLHD 21  
| : : | | | : : |  
Db 381 ARNYKRVAVWMDNYKKVYVD 401

RESULT 15

PCT-US01-01309-222  
; Sequence 222, Application PC/TUS0101309  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc., et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PZ05PCT  
; CURRENT APPLICATION NUMBER: PCT/US01/01309  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 384  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 222.  
; LENGTH: 401  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (37)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (60)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (62)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (66)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (69)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (71)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (73)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (76)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (177)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
PCT-US01-01309-222

Query Match 40.8%; Score 48.5; DB 1; Length 401;  
Best Local Similarity 47.4%; Pred. No. 70;  
Matches 9; Conservative 3; Mismatches 4; Indels 3; Gaps 1;

Qy 5 VSIVEAWSND---DTPYLH 20  
| : : | | | | : : |  
Db 46 VALLEWKRDNRTDDXPXLH 64

Search completed: March 27, 2002, 14:20:27  
Job time: 1574 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:22:48 ; Search time 137.48 seconds  
(without alignments)  
10.540 Million cell updates/sec

Title: US-09-290-049a-14

Perfect score: 119

Sequence: 1 ANNHVSIVEAWSNDTPLYLHD 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 332938 seqs, 68999538 residues

Total number of hits satisfying chosen parameters: 332938

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Pending\_Patents\_AA\_New:\*
- 1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep.\*
  - 2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep.\*
  - 3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep.\*
  - 4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep.\*
  - 5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep.\*
  - 6: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep.\*
  - 7: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep.\*
  - 8: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119	100.0	21	6	US-09-562-328-23
2	119	100.0	21	6	US-09-562-328-25
3	119	100.0	21	6	US-09-290-049A-12
4	119	100.0	21	6	US-09-290-049A-14
5	108	90.8	21	6	US-09-562-328-20
6	108	90.8	21	6	US-09-290-049A-1
7	108	90.8	21	6	US-09-290-049A-10
8	99	83.2	21	6	US-09-562-328-21
9	74	62.2	21	6	US-09-562-328-22
10	74	62.2	21	6	US-09-290-049A-11
11	74	62.2	54.5	6	US-09-604-957-4
12	65	54.6	53.3	6	US-09-604-957-5
13	62	52.1	21	6	US-09-562-328-24
14	62	52.1	21	6	US-09-290-049A-13
15	55	46.2	584	6	US-09-604-957-6
16	50	42.0	1252	6	US-09-614-150-4881
17	48	40.3	535	6	US-09-604-957-7
18	48	40.3	1278	6	US-09-604-957-3
19	46.5	39.1	1604	6	US-09-888-615-73
20	46	38.7	127	6	US-09-675-784A-8751
21	46	38.7	880	6	US-09-614-150-12219
22	45	37.8	199	6	US-09-605-703B-2326
23	45	37.8	638	6	US-09-708-427-25322
24	45	37.8	765	6	US-09-708-427-25321
25	45	37.8	947	6	US-09-708-427-25320

26	45	37.8	972	6	US-09-924-154-16	Sequence 16, Appl
27	44	37.0	15	7	PCT-US02-03987-10190	Sequence 90, Appl
28	43.5	36.6	491	1	PCT-US02-03987-10190	Sequence 10190, A
29	43.5	36.6	491	1	PCT-US02-03987-13756	Sequence 13756, A
30	43.5	36.6	491	6	US-09-815-242-10190	Sequence 10190, A
31	43.5	36.6	491	6	US-09-815-242-13756	Sequence 13756, A
32	43.5	36.6	491	7	US-10-072-851-10190	Sequence 10190, A
33	43.5	36.6	491	7	US-10-072-851-13756	Sequence 13756, A
34	43	36.1	67	7	US-10-050-704-324	Sequence 324, App
35	43	36.1	87	7	US-10-050-704-149	Sequence 149, App
36	43	36.1	89	7	US-10-015-127-12494	Sequence 12494, A
37	43	36.1	304	5	US-09-623-791-83	Sequence 83, Appl
38	43	36.1	434	6	US-09-708-427-12697	Sequence 12697, A
39	43	36.1	461	6	US-09-708-427-12696	Sequence 12696, A
40	43	36.1	473	6	US-09-708-427-12695	Sequence 12695, A
41	43	36.1	2603	6	US-09-897-516-5891	Sequence 5891, Ap
42	42.5	35.7	1059	6	US-09-614-150-1785	Sequence 1785, Ap
43	42	35.3	179	6	US-09-675-784A-8401	Sequence 8401, Ap
44	42	35.3	226	6	US-09-708-427-53175	Sequence 53175, A
45	42	35.3	254	8	US-60-356-051-2856	Sequence 2856, Ap

ALIGNMENTS

RESULT 1  
US-09-562-328-23  
; Sequence 23, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 23  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-23

Query Match 100.0%; Score 119; DB 6; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1.6e-12;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANNHVSIVEAWSNDTPLYLHD 21  
Db 1 ANNHVSIVEAWSNDTPLYLHD 21

RESULT 2  
US-09-562-328-25  
; Sequence 25, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 25  
; LENGTH: 21

```
; TYPE: PRT
; ORGANISM: Streptococcus sp.
US-09-562-328-25

Query Match      100.0%; Score 119; DB 6; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.6e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANNHVSIVEAWSNDNDTPYLHD 21
Db 1 ANNHVSIVEAWSNDNDTPYLHD 21

RESULT 3
US-09-290-049A-12
; Sequence 12, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CARRIES
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290.049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. downei
US-09-290-049A-12

Query Match      100.0%; Score 119; DB 6; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.6e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANNHVSIVEAWSNDNDTPYLHD 21
Db 1 ANNHVSIVEAWSNDNDTPYLHD 21

RESULT 4
US-09-290-049A-14
; Sequence 14, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CARRIES
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290.049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. sobrinus
US-09-290-049A-14

Query Match      100.0%; Score 119; DB 6; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.6e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANNHVSIVEAWSNDNDTPYLHD 21
Db 1 ANNHVSIVEAWSNDNDTPYLHD 21

RESULT 5
US-09-562-328-20
; Sequence 20, Application US/09562328
; GENERAL INFORMATION:
; APPLICANT: LEES, ANDREW
; APPLICANT: TAUBMAN, MARTIN A.
; APPLICANT: SMITH, DANIEL J.
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARRIES
; FILE REFERENCE: 04995.0046-01
; CURRENT APPLICATION NUMBER: US/09/562,328
; CURRENT FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 09/288,965
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Streptococcus sp.
US-09-562-328-20

Query Match      90.8%; Score 108; DB 6; Length 21;
Best Local Similarity 85.7%; Pred. No. 8.8e-11;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANNHVSIVEAWSNDNDTPYLHD 21
Db 1 ANHLSILEAWSNDNDTPYLHD 21

RESULT 6
US-09-290-049A-1
; Sequence 1, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CARRIES
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290.049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: EAW peptide
US-09-290-049A-1

Query Match      90.8%; Score 108; DB 6; Length 21;
Best Local Similarity 85.7%; Pred. No. 8.8e-11;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANNHVSIVEAWSNDNDTPYLHD 21
Db 1 ANHLSILEAWSNDNDTPYLHD 21
```

; LENGTH: 545

FILE REFERENCE: 04995.0046-01

; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-604-957-4

Query Match 62.2%; Score 74; DB 6; Length 545;  
Best Local Similarity 66.7%; Pred. No. 0.00099;  
Matches 14; Conservative 2; Mismatches 5; Indels

Qy	1	ANNHVSIVEAWSNDTPLYLHD	21
Db	75	AINHLSILEAWSNDPOYNKD	95

```

RESULT 12
US-09-604-957-5
Sequence 5, Application US/09604957
GENERAL INFORMATION:
APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
APPLICANT: DIJKHUIZEN, LUBBERT
APPLICANT: RAHAOUI, HAKIM
APPLICANT: LEER, ROBERT-JAN
TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
FILE REFERENCE: BO 43388
CURRENT APPLICATION NUMBER: US/09/604,957
CURRENT FILING DATE: 2000-06-28
PRIOR APPLICATION NUMBER: 00201871.1
PRIOR FILING DATE: 2000-05-25
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn ver. 2.1
SEQ ID NO 5
LENGTH: 523
TYPE: PRT
ORGANISM: Leuconostoc mesenteroides
US-09-604-957-5

```

Query Match 54.6%; Score 65; DB 6; Length 523;  
Best Local Similarity 57.1%; Pred. No. 0.026;  
Matches 12; Conservative 3; Mismatches 6; Indels

QY 1 ANNHVSIVEAWSNDNTPYLHD 21  
||| : ||| : ||| : |  
Db 75 ANQHLSILEDWSHNDPLVTD 95

LT 13  
 09-562-328-24  
 ; Sequence 24, Application US/09562328  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LEES, ANDREW  
 ; APPLICANT: TAUBMAN, MARTIN A.  
 ; APPLICANT: SMITH, DANIEL J.  
 ; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
 ; FILE REFERENCE: 04995.0046-01  
 ; CURRENT APPLICATION NUMBER: US/09/562,328  
 ; CURRENT FILING DATE: 2000-05-01  
 ; PRIOR APPLICATION NUMBER: 09/288,965  
 ; PRIOR FILING DATE: 1999-04-09  
 ; NUMBER OF SEQ ID NOS: 44  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 24  
 ; LENGTH: 21  
 ; TYPE: PRT  
 ; ORGANISM: Streptococcus sp.  
 US-09-562-328-24

```
Query Match          52.1%; Score 62; DB 6; Length 21;
Best Local Similarity 57.9%; Pred. No. 0.0019;
Matches 11: Conservative 4; Mismatches 4; Indels 0; Gaps 0;
```

QY- 1..ANNHVSIVEAWSNDNTPYL 19

```

1 AIDHLSILEAWSGNDNDYV 19
Db

```

```

RESULT 14
US-09-290-049A-13
; Sequence 13, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CARRIES
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290,049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. downei
US-09-290-049A-13

```

Query Match 52.1%; Score 62; DB 6; Length 21;  
Best Local Similarity 57.9%; Pred. NO. 0.0019;  
Matches 11; Conservative 4; Mismatches 4; Indels

Qy 1 ANNHVSIVEAWSNDNTPYL 19  
1 : 1 : 1 : 1 : 1 : 1 :  
pb 1 AIDHLSILEAWSGNDNDYV 19

```

RESULT 15
US-09-604-957-6
; Sequence 6, Application US/09604957
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 584
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
; US-09-604-957-6

```

Query Match 46.2%; Score 55; DB 6; Length 584;  
Best Local Similarity 45.0%; Pred. No. 1.1;  
Matches 9: Conservative 5: Mismatches 6: Indels

Qy 1 ANNHVSIVEAWSNDNTPYLH 20  
|||::|||:| |::  
Db 75 ANKHLSELDWNGKDPOYVN 94

Search completed: March 27, 2002, 14:22:48  
Job time: 1695 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:01:20 ; Search time 102.51 Seconds  
(without alignments)  
15.605 Million cell updates/sec

Title: US-09-290-049a-14  
Perfect score: 119  
Sequence: 1 ANNHVSIVEAWSNDNTPYLHD 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	119	100.0	1592	2 A38175	glucosyltransferas
2	108	90.8	1475	2 B33135	gtfB protein precu
3	99	83.2	1375	2 JT0345	dextranucrase (EC
4	74	62.2	1431	2 A45866	dextranucrase (EC
5	68	57.1	1508	2 T31098	probable dextranu
6	67	56.3	1365	2 A41483	glucosyltransferas
7	61	51.3	1449	2 T30857	glucosyltransferas
8	61	51.3	1449	2 T30852	glucosyltransferas
9	57	47.9	1577	2 T30858	glucosyltransferas
10	56	47.1	1518	2 A44811	glucosyltransferas
11	55	46.2	1599	2 S22737	glucosyltransferas
12	53	44.5	175	2 C86205	hypothetical prote
13	49	41.2	336	1 Q0BE40	BglF2 protein - hu
14	49	41.2	524	2 D82220	conserved hypothet
15	48	40.3	331	2 B48445	glyceraldehide-3-p
16	46	38.7	313	2 S59448	hypothetical prote
17	46	38.7	490	2 H70538	probable ppdk prot
18	45	37.8	337	2 C64233	glyceraldehide-3-p
19	45	37.8	347	2 T48610	hypothetical prote
20	45	37.8	525	2 T40088	RhoGEF domain cont
21	45	37.8	947	2 E86362	hypothetical prote
22	45	37.8	1384	2 T02748	hypothetical prote
23	45	37.8	4848	2 T30289	pristinamycin I sy
24	44.5	37.4	835	2 B64689	site-specific DNA-
25	44	37.0	741	2 JC5142	X-Pro dipeptidyl-p
26	44	37.0	1122	2 S64443	probable membrane
27	44	37.0	1131	2 T16217	hypothetical prote
28	44	37.0	1313	2 T29027	hypothetical prote
29	43.5	36.6	491	1 D64947	glucose-6-phosphat

30 43.5 36.6 491 2 F85797 glucose-6-phosphat  
31 43 36.1 78 2 S76593 hypothetical prote  
32 43 36.1 194 2 H72037 conserved hypothet  
33 43 36.1 194 2 C86586 CT647 hypothetical  
34 43 36.1 275 2 S55978 hypothetical prote  
35 43 36.1 284 2 JC7148 heat-shock sigma f  
36 43 36.1 331 1 DEUTGC glyceraldehide-3-p  
37 43 36.1 346 2 E70715 hypothetical prote  
38 43 36.1 396 2 T35024 probable glutathio  
39 43 36.1 441 2 T36737 probable homogeniti  
40 43 36.1 453 2 T16795 hypothetical prote  
41 43 36.1 473 2 T04218 hypothetical prote  
42 43 36.1 504 2 C86428 hypothetical prote  
43 43 36.1 528 2 E71525 probable ABC trans  
44 43 36.1 603 2 S03020 dopamine beta-mono  
45 43 36.1 899 2 B48586 suppressor of hair

ALIGNMENTS

RESULT 1  
A38175 glucosyltransferase precursor - Streptococcus sobrinus  
C:Species: Streptococcus sobrinus  
C:Date: 28-Aug-1992 #sequence\_revision 28-Aug-1992 #text\_change 15-Oct-1999  
C:Accession: A38175  
R:Abo, H.; Msumura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.  
J. Bacteriol. 173, 989-996, 1991  
A:Title: Peptide sequences for sucrose splitting and glucan binding within Streptococ  
A:Reference number: A38175; MUID:91123227  
A:Accession: A38175  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1592 <ABO>  
A:Cross-references: GB:D90213; NID:g217032; PIDN:BAA14241.1; PID:d1014946; PID:g21703  
C:Superfamily: cpl repeat homology  
F:1093-1113/Domain: cpl repeat homology <CP1>  
F:1222-1241/Domain: cpl repeat homology <CP2>  
F:1287-1306/Domain: cpl repeat homology <CP3>  
F:1330-1351/Domain: cpl repeat homology <CP4>  
F:1352-1371/Domain: cpl repeat homology <CP5>  
F:1402-1420/Domain: cpl repeat homology <CP6>  
F:1465-1484/Domain: cpl repeat homology <CP7>  
F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match 100.0%; Score 119; DB 2; Length 1592;  
Best Local Similarity 100.0%; Pred. No. 4.2e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANNHVSIVEAWSNDNTPYLHD 21  
Db 477 ANNHVSIVEAWSNDNTPYLHD 497

RESULT 2  
B33135 gtfB protein precursor - Streptococcus mutans  
C:Species: Streptococcus mutans  
C:Date: 23-Oct-1990 #sequence\_revision 23-Oct-1990 #text\_change 15-Oct-1999  
C:Accession: B33135; A33128  
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
J. Bacteriol. 169, 4263-4270, 1987  
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.  
A:Reference number: A33135; MUID:87308013  
A:Accession: B33135  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1475 <SHI>  
A:Cross-references: GB:M17361; NID:g153639; PIDN:AAA88588.1; PID:g153640  
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
submitted to the Protein Sequence Database, September 1990

A:Reference number: A33128

A:Accession: A33128

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-171,173-641,'N',643-1475 <SH2>

A:Experimental source: strain GS-5

C:Superfamily: cpl repeat homology

F:1096-1115/Domain: cpl repeat homology <CP1>

F:1224-1243/Domain: cpl repeat homology <CP2>

F:1289-1308/Domain: cpl repeat homology <CP3>

F:1354-1373/Domain: cpl repeat homology <CP4>

F:1419-1438/Domain: cpl repeat homology <CP5>

Query Match

Best Local Similarity 90.8%; Score 108; DB 2; Length 1475;

Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANNHVSIVEAWSNDNTPYLHD 21

Db 481 ANDHLSILEAWSNDNTPYLHD 501

RESULT 3

JT0345

dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)

N:Alternate names: sucrose 6-glucosyltransferase

C:Species: Streptococcus mutans

C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 24-Sep-1999

C:Accession: JT0345; C33135

R:Ueda, S.; Shiroza, T.; Kuramitsu, H.K.

Gene 69, 101-109, 1988

A:Title: Sequence analysis of the gtfc gene from Streptococcus mutans GS-5.

A:Reference number: JT0345; MUID:89137980

A:Accession: JT0345

A:Molecule type: DNA

A:Residues: 1-1375 <UED>

A:Experimental source: GS-5

R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

J. Bacteriol. 169, 4263-4270, 1987

A:Title: Sequence analysis of the gtfb gene from Streptococcus mutans.

A:Reference number: A33135; MUID:87308013

A:Accession: C33135

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-349 <SH2>

A:Cross-references: GB:M17361

A:Genetics:

A:Gene: gtfc

C:Function:

A:Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans

C:Superfamily: cpl repeat homology

C:Keywords: duplication; glycosyltransferase; hexosyltransferase

F:1-34/Domain: signal sequence #status predicted <SIG>

F:35-1375/Product: glucosyltransferase #status predicted <NAT>

F:1126-1145/Domain: cpl repeat homology <CP1>

F:1253-1272/Domain: cpl repeat homology <CP2>

F:1318-1337/Domain: cpl repeat homology <CP3>

Query Match

Best Local Similarity 83.2%; Score 99; DB 2; Length 1375;

Matches 17; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ANNHVSIVEAWSNDNTPYLHD 21

Db 507 ANDHLSILEAWSNDNTPYLHD 527

RESULT 4

A45866

dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans

C:Species: Streptococcus mutans

C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999

C:Accession: A45866

R:Honda, O.; Kato, C.; Kuramitsu, H.K.

J. Gen. Microbiol. 136, 2099-2105, 1990

A:Title: Nucleotide sequence of the Streptococcus mutans gtfd gene encoding the gluco

A:Reference number: A45866; MUID:91100958

A:Accession: A45866

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1431 <HON>

A:Cross-references: GB:M29296

C:Superfamily: cpl repeat homology

C:Keywords: glycosyltransferase; hexosyltransferase

F:181-201/Domain: cpl repeat homology <CP1>

F:1127-1146/Domain: cpl repeat homology <CP2>

F:1192-1211/Domain: cpl repeat homology <CP3>

F:1257-1276/Domain: cpl repeat homology <CP4>

F:1277-1297/Domain: cpl repeat homology <CP5>

F:1321-1340/Domain: cpl repeat homology <CP6>

F:1341-1361/Domain: cpl repeat homology <CP6>

F:1385-1404/Domain: cpl repeat homology <CP7>

Query Match

Best Local Similarity 62.2%; Score 74; DB 2; Length 1431;

Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ANNHVSIVEAWSNDNTPYLHD 21

Db 495 AINHLISLEAWSNDNTPQYKND 515

RESULT 5

T31098

probable dextranucrase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides

C:Species: Leuconostoc mesenteroides

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000

C:Accession: T31098

R:Monchois, V.; Rемаud-Simeon, M.; Monsan, P.; Willemot, R.M.

FEMS Microbiol. Lett. 159, 307-315, 1998

A:Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase

A:Reference number: Z20981; MUID:98164374

A:Accession: T31098

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1508 <NON>

A:Cross-references: EMBL:AF030129; NID:g2766611; PID:g2766612; PIDN:AB95453.1

A:Experimental source: strain NRRL B-1299

C:Genetics:

A:Gene: dsrB

C:Function:

A:Description: produces dextran composed only of alpha(1-6) glucosidic bonds

C:Keywords: glycosyltransferase; hexosyltransferase

Query Match

Best Local Similarity 57.1%; Score 68; DB 2; Length 1508;

Matches 12; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 ANNHVSIVEAWSNDNTPYLHD 21

Db 563 ANQHLSILEAWSNDNTPYVKD 583

RESULT 6

A41483

glucosyltransferase (EC 2.4.1.-) gtfs precursor - Streptococcus sobrinus

C:Species: Streptococcus sobrinus

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 15-Oct-1999

C:Accession: A41483

R:Gillmore, K.S.; Russell, R.R.B.; Ferretti, J.J.

Infect. Immun. 58, 2452-2458, 1990

A:Title: Analysis of the Streptococcus downei gtfs gene, which specifies a glucosyltr

A:Reference number: A41483; MUID:90316665

A:Accession: A41483

A:Molecule type: DNA

A:Residues: 1-1365 <GIL>

A:Cross-references: GB:M30943; NID:g153652; PIDN:AAA26898.1; PID:g153653

C:Genetics:

A:Gene: gtf5

C:Superfamily: cpl repeat homology

C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 56.3%; Score 67; DB 2; Length 1365;

Best Local Similarity 57.1%; Pred. No. 0.044;

Matches 12; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 ANNHVSIVEAWSNDNTPYLHD 21

I:|:|:|:|:|:|:|:|:|:|:|:|:|

Db 467 AIDHLSILEAWSGNDNDYVKD 487

RESULT 7

T30857

glucosyltransferase - Streptococcus salivarius

C:Species: Streptococcus salivarius

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999

C:Accession: T30857

R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.

Infect. Immun. 63, 609-621, 1995

A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pr

A:Reference number: Z20909; MUID:95122197

A:Accession: T30857

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1449 <SIM>

A:Cross-references: EMBL:L35495; NID:g662378; PID:g662379; PIDN:AAC41412.1

C:Genetics:

A:Gene: gtf1

Query Match 51.3%; Score 61; DB 2; Length 1449;

Best Local Similarity 57.1%; Pred. No. 0.41;

Matches 12; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 ANNHVSIVEAWSNDNTPYLHD 21

I:|:|:|:|:|:|:|:|:|:|:|:|:|

Db 536 AIKHLISILEAWSNDAYNYED 556

RESULT 8

T30852

glucosyltransferase N - Streptococcus salivarius (fragment)

C:Species: Streptococcus salivarius

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999

C:Accession: T30852

R:Jaffe, R.I.

submitted to the EMBL Data Library, February 1998

A:Description: Streptococcus salivarius V1477 gtfN.

A:Reference number: Z20854

A:Accession: T30852

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1449 <JAF>

A:Cross-references: EMBL:AF049609; NID:g2935545; PID:g2935546; PIDN:AAC05156.1

C:Genetics:

A:Gene: gtfN

Query Match 51.3%; Score 61; DB 2; Length 1449;

Best Local Similarity 57.1%; Pred. No. 0.41;

Matches 12; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 ANNHVSIVEAWSNDNTPYLHD 21

I:|:|:|:|:|:|:|:|:|:|:|:|:|

Db 536 AIKHLISILEAWSNDAYNYED 556

RESULT 9

T30858

glucosyltransferase - Streptococcus salivarius

C:Species: Streptococcus salivarius

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999

C:Accession: T30858

R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.

Infect. Immun. 63, 609-621, 1995

A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for

A:Reference number: Z20909; MUID:95122197

A:Accession: T30858

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1577 <SIM>

A:Cross-references: EMBL:L35928; NID:g662380; PID:g662381; PIDN:AAC41413.1

C:Genetics:

A:Gene: gtfm

Query Match 47.9%; Score 57; DB 2; Length 1577;

Best Local Similarity 61.1%; Pred. No. 1.9;

Matches 11; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 HVSIVEAWSNDNTPYLHD 21

I:|:|:|:|:|:|:|:|:|:|:|:|:|

Db 591 HLSILEAWSYNDHQYNKD 608

RESULT 10

A44811

glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius

C:Species: Streptococcus salivarius

C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 15-Oct-1999

C:Accession: A44811; S22726; S28809

R:Giffard, P.M.; Simpson, C.L.; Millward, C.P.; Jacques, N.A.

J. Gen. Microbiol. 137, 2577-2593, 1991

A:Title: Molecular characterization of a cluster of at least two glucosyltransferase

A:Reference number: A44811; MUID:92148377

A:Accession: A44811

A:Molecule type: DNA

A:Residues: 1-1518 <GIF>

A:Cross-references: EMBL:Z11873; NID:g47526; PIDN:CAA77900.1; PID:g47527

A:Note: sequence extracted from NCBI backbone (NCBIN:81050, NCBIP:81052)

C:Genetics:

A:Gene: gtfJ

C:Superfamily: cpl repeat homology

C:Keywords: glycosyltransferase; hexosyltransferase

F:1307-1326/Domain: cpl repeat homology <CP4>

Query Match 47.1%; Score 56; DB 2; Length 1518;

Best Local Similarity 60.0%; Pred. No. 2.6;

Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 HVSIVEAWSNDNTPY 18

I:|:|:|:|:|:|:|:|:|:|:|:|:|

Db 504 HISVLEAWSLNDNH 518

RESULT 11

S22737

glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius

C:Species: Streptococcus salivarius

C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 21-Jan-2000

C:Accession: S22737; S28810; B44811; S22727

R:Jacques, N.

submitted to the EMBL Data Library, March 1992

A:Reference number: S22726

A:Accession: S22737

A:Molecule type: DNA

A:Residues: 1-1599 <JAC>  
A:Cross-references: EMBL:Z11872; NID:g47530; PIDN:CAA77898.1; PID:g47531  
A:Experimental source: ATCC 25975  
R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.  
J. Gen. Microbiol. 137, 2577-2593, 1991  
A:Title: Molecular characterization of a cluster of at least two glucosyltransferase genes  
A:Reference number: A44811; MUID:92148377  
A:Accession: S28810  
A:Molecule type: DNA  
A:Residues: 1-51 <GIE>  
A:Cross-references: EMBL:Z11873  
C:Genetics:  
C:Superfamily: cpl repeat homology  
C:Keywords: glucosyltransferase; hexosyltransferase  
F:1456-1475/Domain: cpl repeat homology <CPR>

Query Match 46.2%; Score 55; DB 2; Length 1599;  
Best Local Similarity 50.0%; Pred. No. 4;  
Matches 9; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 4 HVSIVEAWSNDTPYLHD 21  
::||:|||| || |:::  
Db 494 NISILEAWSNDPYVNE 511

RESULT 12  
C86205  
hypothetical protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: C86205  
R:Thellogis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: C86205  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-175 <STO>  
A:Cross-references: GB:AE005172; NID:g8954041; PIDN:AAF82215.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 44.5%; Score 53; DB 2; Length 175;  
Best Local Similarity 42.1%; Pred. No. 0.59;  
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 3 NHSVIVEAWSNDTPYLHD 21  
|| ::|||: | |  
Db 109 NHQVEIDAWSHQKPLWTD 127

RESULT 13  
QB8E40  
BGLF2 protein - human herpesvirus 4 (strain B95-8)  
C:Species: human herpesvirus 4, Epstein-Barr virus  
C:Date: 25-Feb-1985 #sequence\_revision 25-Feb-1985 #text\_change 16-Jul-1999  
C:Accession: C43044; J01381; A03784; A03794; S33036  
R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.  
Mol. Biol. Med. 1, 21-45, 1983  
A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus  
A:Reference number: A93065; MUID:85035713  
A:Accession: C43044

A:Molecule type: DNA  
A:Residues: 1-336 <BAN>  
A:Cross-references: EMBL:V01555; NID:g59074; PIDN:CAA24831.1; PID:g1334895  
R:Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.  
Nature 310, 207-211, 1984  
A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.  
A:Reference number: A03794; MUID:84270667  
A:Contents: annotation; protein coding region  
A:Note: neither amino acid nor nucleotide sequence is given  
R:Chen, M.R.; Hsu, T.Y.; Lin, S.W.; Chen, J.Y.; Yang, C.S.  
J. Gen. Virol. 72, 3047-3055, 1991  
A:Title: Cloning and characterization of cDNA clones corresponding to transcripts fro  
A:Reference number: J01381; MUID:92113548  
A:Accession: J01381  
A:Molecule type: mRNA  
A:Residues: 1-336 <CHE>  
A:Cross-references: GB:S77132; NID:g243314; PIDN:AAB21113.1; PID:g243315  
C:Superfamily: Epstein-Barr virus BGLF2 protein

Query Match 41.2%; Score 49; DB 1; Length 336;  
Best Local Similarity 41.2%; Pred. No. 5.4;  
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 1 ANHVSIVEAWSNDTP 17  
| ||:|: |::|:  
Db 216 AGAHVILRGWTEDDSP 232

RESULT 14  
D82220  
conserved hypothetical protein VC1268 [imported] - Vibrio cholerae (strain N16961 ser  
C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C:Accession: D82220  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.  
ardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers  
L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833  
A:Accession: D82220  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-524 <HEI>  
A:Cross-references: GB:AE004206; GB:AE003852; NID:g9655749; PIDN:AAF94427.1; GSPDB:GN  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:  
A:Map position: 1

Query Match 41.2%; Score 49; DB 2; Length 524;  
Best Local Similarity 43.5%; Pred. No. 9.1;  
Matches 10; Conservative 4; Mismatches 7; Indels 2; Gaps 1;

QY 1 ANHVSIVEAWSNDTP--PYLHD 21  
||:|: |::| | | |  
Db 365 ANHMEIRGWSDEVIDPALID 387

RESULT 15  
B48445  
glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12) - Leishmania mexicana  
C:Species: Leishmania mexicana  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 11-Jun-1999  
C:Accession: B48445; S25142  
R:Hannart, V.; Blaauw, M.; Kohl, L.; Allert, S.; Oppendoes, F.R.; Michels, P.A.M.  
Mol. Biochem. Parasitol. 55, 115-126, 1992  
A:Title: Molecular analysis of the cytosolic and glycosomal glyceraldehyde-3-phosphat  
A:Reference number: A48445; MUID:93063042  
A:Accession: B48445  
A>Status: preliminary

A: Molecule type: DNA  
A: Residues: 1-331 <HAN>  
A: Cross-references: EMBL:X65220; NID:g9552; PIDN:CAA46323.1; PID:g9553  
C: Superfamily: glyceraldehyde-3-phosphate dehydrogenase  
C: keywords: oxidoreductase

Query Match 40.3%; Score 48; DB 2; Length 331;  
Best Local Similarity 42.1%; Pred. No. 7.6;  
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 NNHVSIVEAWSNDTPYLH 20  
|:| :| ||| | |  
Db 301 NDHFVKLVSWYDNETGYSH 319

Search completed: March 27, 2002, 14:01:21  
Job time: 483 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:26:05 ; Search time 188.53 Seconds  
(without alignments)  
16.293 Million cell updates/sec

Title: US-09-290-049A-13  
Perfect score: 113  
Sequence: 1 AIDHLSILEAWSGNDNDYVKQ 21

Scoring table: BLOSUM62  
Gapop 10.0 , Capext 0.5

Searched: 473505 seqs, 146272329 residues  
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL\_17:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	95.6	1338	2 Q9WXJ4	Q9wxj4 streptococc
2	79	69.9	1575	2 Q9LCH3	Q9lch3 streptococc
3	77	68.1	1577	2 Q54178	Q54178 streptococc
4	73	64.6	1390	2 Q69385	Q69385 streptococc
5	73	64.6	1455	2 Q69391	Q69391 streptococc
6	73	64.6	1577	2 Q55265	Q55265 streptococc
7	71	62.8	1455	2 Q69382	Q69382 streptococc
8	71	62.8	1455	2 Q69388	Q69388 streptococc
9	71	62.8	1455	2 Q69397	Q69397 streptococc
10	69	61.1	1477	2 Q9LA66	Q9la66 leuconostoc
11	69	61.1	1508	2 Q52224	Q52224 leuconostoc
12	69	61.1	1508	2 Q9E2H5	Q9e2h5 leuconostoc
13	69	61.1	1512	2 Q9WXJ5	Q9wxj5 streptococc
14	69	61.1	1518	2 Q00600	Q00600 streptococc
15	67	59.3	1449	2 Q68542	Q68542 streptococc
16	67	59.3	1449	2 Q55264	Q55264 streptococc
17	66	58.4	1016	2 Q9LCJ7	Q9lcj7 leuconostoc
18	65	57.5	2057	2 Q9RE05	Q9re05 leuconostoc
19	62	54.9	1590	2 Q55263	Q55263 streptococc

20	62	54.9	1590	2 Q59983	Q59983 streptococc
21	61	54.0	1599	2 Q00599	Q00599 streptococc
22	58	51.3	1527	2 Q9ZARA	Q9zara leuconostoc
23	49	43.4	232	1 Q9HSM9	Q9hsm9 halobacteri
24	48	42.5	378	10 Q9SZH3	Q9szh3 arabidopsis
25	48	42.5	695	2 Q9Z9G1	Q9z9g1 chlamydia p
26	48	42.5	704	2 Q9K1Z6	Q9k1z6 chlamydia p
27	48	42.5	3198	12 Q9IW34	Q9iw34 pea seed-bo
28	47	41.6	298	5 Q19058	Q19058 caenorhabdi
29	47	41.6	769	10 Q9LNU1	Q9lnul arabidopsis
30	47	41.6	3199	12 Q85074	Q85074 pea seed-bo
31	46	40.7	255	11 Q9CXL4	Q9cxl4 mus musculu
32	46	40.7	287	2 Q9RK11	Q9rk11 streptomyce
33	46	40.7	458	10 Q9LMT4	Q9lmt4 arabidopsis
34	46	40.7	514	2 Q9A4Q6	Q9a4q6 caulobacter
35	46	40.7	536	9 Q38324	Q38324 lactococcus
36	46	40.7	1802	10 Q48647	Q48647 oryza sativ
37	45.5	40.3	722	4 Q9Y223	Q9y223 homo sapien
38	45.5	40.3	722	11 Q35826	Q35826 rattus norv
39	45.5	40.3	722	11 Q9Z0P6	Q9z0p6 mus musculu
40	45	39.8	409	5 Q9W0S3	Q9w0s3 drosophila
41	45	39.8	899	2 P72473	P72473 serratia ma
42	45	39.8	2367	2 Q46034	Q46034 clostridium
43	45	39.8	2367	2 Q9F931	Q9f931 clostridium
44	44.5	39.4	907	5 Q9NE48	Q9ne48 leishmania
45	44	38.9	389	12 Q84432	Q84432 paramecium

ALIGNMENTS

RESULT 1  
Q9WXJ4  
ID Q9WXJ4 PRELIMINARY: PRT: 1338 AA.  
AC Q9WXJ4;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE GTF-S.  
GN GTF-S.  
OS Streptococcus criceti.  
OG Plasmid pAM1.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1333;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HS-6;  
RA Inoue M., Fukui K., Miyagi A.;  
RT "S.cricetus glucosyltransferase(gtfs and gtft) genes.";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AB026123; BAA77236.1; -;  
DR HSSP; P06278; IUVS.  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 10.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
DR Plasmid.  
SQ SEQUENCE 1338 AA; 148558 MW; 0A90C8E10E15D99B CRC64;

Query Match 95.6%; Score 108; DB 2; Length 1338;  
Best Local Similarity 100.0%; Pred. No. 1.6e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AIDHLSILEAWSGNDNDYVK 20  
|||||  
Db 437 AIDHLSILEAWSGNDNDYVK 456

RESULT 2  
Q9LCH3  
ID Q9LCH3 PRELIMINARY: PRT: 1575 AA.

```

SQ  SEQUENCE      1577 AA;  177805 MW;  5AE0328DC5E08D18 CRC64;

Query Match          68.1%;  Score 77;  DB 2;  Length 1577;
Best Local Similarity 75.0%;  Pred No. 0.0017;
Matches 15;  Conservative 1;  Mismatches 4;  Indels 0;  Gaps 0;

QY  1 AIDHLSILEAWSGNDNDYK 20
    I: ||||| ||| I
Db   548 ALKHLISLEAWSNDPDYK 567

RESULT 4
O69385 PRELIMINARY; PRT; 1390 AA.
ID O69385 AC O69385;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DE GLUCOSYLTRANSFERASE-SI.
GN GTFC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT4245;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D88655; BAA26106.1; -.
DR InterPro; IPR002479; CW_binding.
DR DR PFam; PF01473; CW_binding.1; 7.
DR DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE      1390 AA;  155375 MW;  8847E4956EF05E9F CRC64;

Query Match          64.6%;  Score 73;  DB 2;  Length 1390;
Best Local Similarity 73.7%;  Pred No. 0.0064;
Matches 14;  Conservative 1;  Mismatches 4;  Indels 0;  Gaps 0;

QY  1 AIDHLSILEAWSGNDNDYK 19
    I: ||||| ||| I
Db   507 ANDHLSILEAWSNDTPYL 525

RESULT 5
O69391 PRELIMINARY; PRT; 1455 AA.
ID O69391 AC O69391;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DE GLUCOSYLTRANSFERASE-SI.
GN GTFC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT4251;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT streptococcus mutans.";

```



DE GLUCOSYLTRANSFERASE-SI.  
GN GTEC.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MT4467;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
RA Kimura S., Hamada S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of  
RT Streptococcus mutans";  
RL FEMS Microbiol. Lett. 161:331-336(1998).  
DR EMBL; D89978; BAA26120.1; .  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
PFam; PF01473; CW\_binding\_1; 9.  
PFam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1455 AA; 162913 MW; A1263427BF24E8E1 CRC64;  
Query Match 62.8%; Score 71; DB 2; Length 1455;  
Best Local Similarity 73.7%; Pred. No. 0.014;  
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 AIDHLSILEAWSGNDNDYV 19  
Db 507 ANDHLSILEAWSYNDTPYL 525  
RESULT 10  
Q9L466  
ID Q9L466 PRELIMINARY; PRT; 1477 AA.  
AC Q9L466;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE DEXTRANSUCRASE (EC 2.4.1.5).  
GN DSRC.  
OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OC Leuconostoc.  
OX NCBI\_TaxID=1245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NRRL B-1355;  
RA Arguello-Morales M.A., Remaud-Simeon M., Pizzut S., Sarcabal P.,  
RA Willemot R.M., Monsan P.;  
RT "Sequence analysis of the gene encoding alternansucrase, a sucrose  
RT glucosyltransferase from Leuconostoc mesenteroides NRRL B-1355.";  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ250172; CAB76565.1; .  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 14.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase; Glycosyltransferase.  
SQ SEQUENCE 1477 AA; 164886 MW; E6F5710DEDFCB831 CRC64;  
Query Match 61.1%; Score 69; DB 2; Length 1477;  
Best Local Similarity 70.0%; Pred. No. 0.03;  
Matches 14; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 AIDHLSILEAWSGNDNDYV 20  
Db 532 ANQHLISILEDWSHNDPEYVK 551  
RESULT 11

Q52224  
ID Q52224 PRELIMINARY; PRT; 1508 AA.  
AC Q52224;  
DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE 6-  
DE GLUCOSYLTRANSFERASE).  
GN DSRB.  
OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OC Leuconostoc.  
OX NCBI\_TaxID=1245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NRRL B-1299;  
RA Monchois V., Remaud-Simeon M., Monsan P., Willemot R.M.;  
RA FEMS Microbiol. Lett. 0:0-0(1998).  
CC -I- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) = D-  
CC FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).  
DR EMBL; AF030129; AAB95453.1; .  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 14.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase; Glycosyltransferase.  
SQ SEQUENCE 1508 AA; 168511 MW; E70CECB57A70D1F0 CRC64;  
Query Match 61.1%; Score 69; DB 2; Length 1508;  
Best Local Similarity 70.0%; Pred. No. 0.03;  
Matches 14; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 AIDHLSILEAWSGNDNDYV 20  
Db 563 ANQHLISILEDWSHNDPEYVK 582  
RESULT 12  
Q9EZH5  
ID Q9EZH5 PRELIMINARY; PRT; 1508 AA.  
AC Q9EZH5;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE DEXTRANSUCRASE DSRB742.  
GN DSRB742.  
OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OC Leuconostoc.  
OX NCBI\_TaxID=1245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B-742CB;  
RA Kim H.-S., Kim D., Ryu H.-J., Robyt J.F.;  
RT "Leuconostoc mesenteroides B-742CB, a dextranucrase gene";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF294469; AAC38021.1; .  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 14.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
SQ SEQUENCE 1508 AA; 168542 MW; E2FCFA0F87AE4F3A CRC64;  
Query Match 61.1%; Score 69; DB 2; Length 1508;  
Best Local Similarity 70.0%; Pred. No. 0.03;  
Matches 14; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 AIDHLSILEAWSGNDNDYV 20  
Db 563 ANQHLISILEDWSHNDPEYVK 582

```
RESULT 13
Q9WXJ5
ID Q9WXJ5 PRELIMINARY; PRT; 1512 AA.
AC Q9WXJ5;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GTF-S.
GN GTF.
OS Streptococcus criceti.
OC Plasmid pAMI.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1333;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HS-6;
RA Inoue M., Fukui K., Miyagi A.;
RT "S.cricetus glucosyltransferase(gtfs and gtf) genes.";
Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
EMBL; AB026123; BAA77237.1; -.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 14.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Plasmid.
SQ SEQUENCE 1512 AA; 167145 MW; 4C03D9CBC601FC14 CRC64;

Query Match 61.1%; Score 69; DB 2; Length 1512;
Best Local Similarity 72.2%; Pred. No. 0.03;
Matches 13; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AIDHLSILEAWSGNDNDY 18
I: I:||||| I I I
Db 485 ALAHISILEAWSYNDNY 502

RESULT 14
Q00600
ID Q00600 PRELIMINARY; PRT; 1518 AA.
AC Q00600;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE I (EC 2.4.1.5) (GTF) (DEXTRANSUCRASE).
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTFJ.
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25975;
RA Giffard P.M., Simpson C.L., Milward C.P., Jacques N.A.;
RT "Molecular characterization of a cluster of at least two glucosyltransferase genes in Streptococcus salivarius ATCC 25975.";
J. Gen. Microbiol. 137:2577-2593(1991).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) = D-FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- DISEASE: DENTAL CARIES.
DR EMBL; Z11873; CAA77900.1; -.
DR EMBL; M64111; AAA26896.1; -.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
```

```
DR Pfam; PF01473; CW_binding_1; 13.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase; Repeat; Dental caries.
FT DOMAIN 1307 1482 6 DIRECT REPEATS.
FT REPEAT 1307 1338 PEPEAT 1.
FT REPEAT 1339 1352 PEPEAT 2.
FT REPEAT 1372 1403 PEPEAT 3.
FT REPEAT 1404 1417 PEPEAT 4.
FT REPEAT 1437 1468 PEPEAT 5.
FT REPEAT 1469 1482 PEPEAT 6.
SQ SEQUENCE 1518 AA; 167730 MW; DAA41F717098B59A CRC64;
```

Query Match 61.1%; Score 69; DB 2; Length 1518;  
Best Local Similarity 66.7%; Pred. No. 0.03;  
Matches 12; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AIDHLSILEAWSGNDNDY 18  
I: I:||||| I I I  
Db 501 ALAHISVLEAWSLNDNH 518

```
RESULT 15
O68542
ID O68542 PRELIMINARY; PRT; 1449 AA.
AC O68542;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE N (FRAGMENT).
GN GTFN.
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V1477;
RA Jaffe R.I.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF049609; AAC05156.1; -.
DR InterPro; IPR002479; CW_binding.
DR Pfam; PF01473; CW_binding_1; 8.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
FT NON_TER 1449 1449
SQ SEQUENCE 1449 AA; 159895 MW; 0700F6D748471BFB CRC64;
```

Query Match 59.3%; Score 67; DB 2; Length 1449;  
Best Local Similarity 77.8%; Pred. No. 0.06;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AIDHLSILEAWSGNDNDY 18  
I: I:||||| I I I  
Db 536 AIKHUSILEAWSNDAY 553

Search completed: March 27, 2002, 14:26:06  
Job time: 1678 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:57:57 ; Search time 198.55 Seconds  
(without alignments)  
7.834 Million cell updates/sec

Title: US-09-290-049A-14

Perfect score: 119

Sequence: 1 ANNHVSIWEASDNDTPYLHD 21

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_1101.\*

- 1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.\*
- 2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.\*
- 3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.\*
- 4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.\*
- 5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.\*
- 6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.\*
- 7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.\*
- 8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.\*
- 9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.\*
- 10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.\*
- 11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.\*
- 12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.\*
- 13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.\*
- 14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.\*
- 15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.\*
- 16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.\*
- 17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.\*
- 18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.\*
- 19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.\*
- 20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.\*
- 21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.\*
- 22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	119	100.0	1592	14 AAR32925	Glucosyltransferase
2	57	47.9	1577	17 AAB10667	Alpha-D-glucosyltr
3	55	46.2	2057	21 AAB10667	L. mesenteroides a
4	53	44.5	195	21 AAG09941	Arabidopsis thalia
5	51	42.9	1195	20 AAU75420	T. thermophilus nit
6	48.5	40.8	401	22 AAM43544	Human polypeptide
7	46	38.7	486	22 AAB36552	Lawsonia intracell
8	45	37.8	174	21 AAG60587	Arabidopsis thalia
9	45	37.8	185	21 AAG60586	Arabidopsis thalia
10	45	37.8	189	21 AAG60585	Arabidopsis thalia
11	45	37.8	199	22 AAG90817	C glutamicum prote

12	45	37.8	305	21 AAG09845	Arabidopsis thalia
13	45	37.8	339	21 AAG09844	Arabidopsis thalia
14	45	37.8	603	18 AAW17889	Photobacterium lum
15	45	37.8	1095	18 AAW17888	Photobacterium lum
16	45	37.8	1366	21 AAB41789	Human ORFX ORF1553
17	43.5	36.6	202	16 AAR70641	Morchella sp. alph
18	43.5	36.6	1092	17 AAR91234	Algal alpha-1,4-gl
19	43.5	36.6	1092	20 AAW88257	Gracilariaopsis lem
20	43	36.1	88	21 AAB38373	Human secreted pro
21	43	36.1	151	20 AAY35397	Chlamydia pneumoni
22	43	36.1	160	22 AAE00334	Human membrane-bou
23	43	36.1	163	22 AAE00333	Human membrane-bou
24	43	36.1	220	22 AAG82161	S. epidermidis ope
25	43	36.1	253	22 AAE00332	Human membrane-bou
26	43	36.1	256	22 AAE00331	Human membrane-bou
27	43	36.1	271	21 AAB18985	Amino acid sequenc
28	43	36.1	271	22 AAE00330	Human membrane-bou
29	43	36.1	304	20 AAY48244	Human prostate can
30	43	36.1	437	21 AAY94930	Human secreted pro
31	43	36.1	541	20 AAY36832	Chlamydia trachoma
32	42	35.3	146	18 AAW27957	Staphylococcus aur
33	42	35.3	201	21 AAW90853	Human ADAM protein
34	42	35.3	334	11 AAR03211	Amino acid sequenc
35	42	35.3	455	22 AAB68328	Amino acid sequenc
36	42	35.3	530	18 AAW14772	Human metalloprote
37	42	35.3	540	21 AAW90851	Human ADAM protein
38	42	35.3	540	21 AAW90864	Human ADAM protein
39	42	35.3	775	21 AAW90855	Human ADAM protein
40	42	35.3	775	21 AAW90865	Human ADAM protein
41	42	35.3	1627	12 AAR12789	M. pneumoniae P1 c
42	42	35.3	1627	15 AAR47911	Mycoplasma pneumon
43	42	35.3	1627	16 AAR67538	Cytadhesin P1. My
44	42	35.3	3898	12 AAR10473	Hog cholera virus
45	41.5	34.9	325	22 AAM41942	Human polypeptide

ALIGNMENTS

RESULT 1  
AAR32925  
ID AAR32925 standard; Protein; 1592 AA.  
XX  
AC AAR32925;  
XX  
DT 28-JUN-1993 (first entry)  
XX  
DE Glucosyltransferase I.  
XX  
KW GT-1; Streptococcus; dental; caries.  
XX  
OS Streptococcus sobrinus.

XX JP05023188-A.  
XX  
PD 02-FEB-1993.  
XX  
PF 25-JUL-1991; 91JP-0186592.  
XX  
PR 25-JUL-1991; 91JP-0186592.  
XX  
PA (FUKUI/) FUKUI I.  
XX  
PA (KATO/) KATO K.  
XX  
DR WPI; 1993-079449/10.  
DR N-ESDB; AAO37760.  
XX  
PT DNA sequence glucosyl:transferase-I - comprises Streptococcus  
PT sobrinus DNA sequence with at least one nucleotide added or  
PT deleted  
XX  
PS Claim 13; Page 15; 29pp; Japanese.  
XX







```
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150586.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160788.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 44.5%; Score 53; DB 21; Length 195;
Best Local Similarity 42.1%; Pred. No. 2.2;
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Oy 3 NHVSIVEAWSNDTPYLHD 21
Db 109 nhqvidawsdqkplwtd 127

RESULT 5
KW Human; antiarthritic; antirheumatic; antiproliferative; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antiulcer; vulnary; anticonvulsant; antibacterial;
```

```
AAW75420
ID AAW75420 standard; protein; 1195 AA.
XX
AC AAW75420;
XX
DT 16-MAR-1999 (first entry)
XX
DE T.thermophilus nitrate reductase alpha subunit.
XX
KW Heat-stable; nitrate reductase; temperature; detection; food; toxicity;
KW carcinogen.
XX
OS Thermus thermophilus.
XX
FH Key Location/Qualifiers
FT Misc-difference 630 /label= unknown
FT Misc-difference 669 /label= unknown
FT Misc-difference 691 /label= unknown
FT Misc-difference /label= unknown
XX
PN ES2121561-A1.
XX
XX 16-NOV-1998.
XX
PF 09-MAY-1997; 97ES-0001003.
XX
PR 09-MAY-1997; 97ES-0001003.
XX
PA (UYMA-) UNIV AUTONOMA MADRID.
XX
DR WPI; 1999-001909/01.
XX
XX Heat stable nitrate reductase for high temperature nitrate detection
PT - comprises Thermus thermophilus derivative enhancing nitrite or
PT nitrate reduction
XX
PS Disclosure; Fig 2; 8pp; Spanish.
XX
CC This sequence represents the amino acid sequence of the Thermus
CC thermophilus heat-stable nitrate reductase alpha subunit. Heat stable
CC nitrate reductase can be used for high-temperature detection of nitrates
CC in samples, e.g. in food, where high levels of nitrates can be toxic or
CC carcinogenic.
XX
SQ Sequence 1195 AA;

Query Match 42.9%; Score 51; DB 20; Length 1195;
Best Local Similarity 42.9%; Pred. No. 37;
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Oy 1 ANHVSIVEAWSNDTPYLHD 21
Db 293 avnhvlkefyadrevpyfqd 313

RESULT 6
AAW43544
ID AAW43544 standard; Protein; 401 AA.
XX
AC AAW43544;
XX
XX 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 222.
XX
KW Human; antiarthritic; antirheumatic; antiproliferative; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antiulcer; vulnary; anticonvulsant; antibacterial;
```

KW antiparasitic; cardiant; gene therapy; cancer; immune disorder;  
KW cardiovascular disorder; neurological disease; infection; human.  
XX Homo sapiens.

XX WO200155308-A2.  
XX 02-AUG-2001.  
XX 17-JAN-2001; 2001WO-US01309.  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189874.  
XX 17-MAR-2000; 2000US-0190076.  
XX 18-APR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216647.  
XX 07-JUL-2000; 2000US-0216880.  
XX 11-JUL-2000; 2000US-0217487.  
XX 11-JUL-2000; 2000US-0217496.  
XX 14-JUL-2000; 2000US-0218290.  
XX 26-JUL-2000; 2000US-0220963.  
XX 26-JUL-2000; 2000US-0220964.  
XX 14-AUG-2000; 2000US-0224518.  
XX 14-AUG-2000; 2000US-0224519.  
XX 14-AUG-2000; 2000US-0225213.  
XX 14-AUG-2000; 2000US-0225214.  
XX 14-AUG-2000; 2000US-0225266.  
XX 14-AUG-2000; 2000US-0225267.  
XX 14-AUG-2000; 2000US-0225268.  
XX 14-AUG-2000; 2000US-0225270.  
XX 14-AUG-2000; 2000US-0225447.  
XX 14-AUG-2000; 2000US-0225757.  
XX 14-AUG-2000; 2000US-0225758.  
XX 14-AUG-2000; 2000US-0225759.  
XX 18-AUG-2000; 2000US-0226279.  
XX 22-AUG-2000; 2000US-0226681.  
XX 22-AUG-2000; 2000US-0226686.  
XX 22-AUG-2000; 2000US-0227182.  
XX 23-AUG-2000; 2000US-0227009.  
XX 30-AUG-2000; 2000US-0228924.  
XX 01-SEP-2000; 2000US-0229287.  
XX 01-SEP-2000; 2000US-0229343.  
XX 01-SEP-2000; 2000US-0229344.  
XX 01-SEP-2000; 2000US-0229345.  
XX 05-SEP-2000; 2000US-0229509.  
XX 05-SEP-2000; 2000US-0229513.  
XX 06-SEP-2000; 2000US-0230437.  
XX 06-SEP-2000; 2000US-0230438.  
XX 08-SEP-2000; 2000US-0231242.  
XX 08-SEP-2000; 2000US-0231243.  
XX 08-SEP-2000; 2000US-0231244.  
XX 08-SEP-2000; 2000US-0231413.  
XX 08-SEP-2000; 2000US-0231414.  
XX 08-SEP-2000; 2000US-0232080.  
XX 08-SEP-2000; 2000US-0232081.  
XX 12-SEP-2000; 2000US-0231968.  
XX 14-SEP-2000; 2000US-0232397.  
XX 14-SEP-2000; 2000US-0232398.  
XX 14-SEP-2000; 2000US-0232399.  
XX 14-SEP-2000; 2000US-0232400.  
XX 14-SEP-2000; 2000US-0232401.  
XX 14-SEP-2000; 2000US-0233063.  
XX 14-SEP-2000; 2000US-0233064.  
XX 14-SEP-2000; 2000US-0233065.  
XX 21-SEP-2000; 2000US-0234223.  
XX 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249246.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251889.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.

PR 05-JAN-2001; 2001US-0259678.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Barash SC, Ruben SM;  
 XX WPI; 2001-488781/53.  
 DR N-PSDB; AAI63850.  
 XX  
 PT New isolated nucleic acids and polypeptides, useful for diagnosing,  
 PT treating and/or preventing human diseases and disorders -  
 XX  
 PS Claim 11; SEQ ID NO 222; 664pp + Sequence Listing; English.  
 XX  
 CC The invention relates to human polynucleotides (AAI63803-AAI64012) and  
 CC the encoded proteins (AAM34497-AAM3660) useful for preventing, treating  
 CC or ameliorating medical conditions e.g. by protein or gene therapy. The  
 CC genes were isolated from a range of human tissues disclosed in the  
 CC specification. The nucleic acids, proteins, antibodies and (ant)agonists  
 CC are useful in the diagnosis, treatment and prevention of: (a) cancer,  
 CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,  
 CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or  
 CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,  
 CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,  
 CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 401 AA;

Query Match 40.8%; Score 48.5; DB 22; Length 401;  
 Best Local Similarity 47.4%; Pred. No. 25;  
 Matches 9; Conservative 3; Mismatches 4; Indels 3; Gaps 1;

OY 5 VSIVEAWSNDN---DRPYLH 20  
 I::: I I I I I I I I  
 Db 46 vallevwkdndtdxpxlh 64

RESULT 7  
 36552  
 AAB36552 standard; Protein: 486 AA.

AC AAB36552;  
 DT 07-MAR-2001 (first entry)  
 XX  
 DE Lawsonia intracellularis flagellar hook protein FlgE SEQ ID NO:1.  
 XX  
 KW Lawsonia intracellularis; flgE; flagellar hook protein; vaccine;  
 KW intestinal disease; immunogenic; diagnosis; antibacterial; swine;  
 KW pig; infection; detection; identification.  
 XX

OS Lawsonia intracellularis.  
 XX  
 PN WO200069904-A1.  
 XX  
 PD 23-NOV-2000.  
 XX  
 PF 11-MAY-2000; 2000WO-AU00437.  
 XX  
 PR 13-MAY-1999; 99US-0133973.  
 XX

XX (PRIZ ) PRIZER PROD INC.  
 PA (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.  
 PA (PIGR-) PIG RES & DEV CORP.  
 XX

PI Panaccio M, Rosey EL, Sinistaj M, Hasse D, Parsons J;  
 PI Ankenbauer RG;  
 XX  
 DR WPI; 2001-016210/02.  
 DR N-PSDB; AAC88037.  
 XX  
 PT New immunogenic Lawsonia FlgE peptide, its nucleic acid and antibody,  
 PT useful in vaccines and diagnosis of Lawsonia infections, particularly  
 PT in swine -  
 XX  
 PS Claim 13; Page 87-90; 97pp; English.  
 XX  
 CC The present sequence is the Lawsonia intracellularis flagellar hook  
 CC protein FlgE. The present invention describes an isolated or recombinant  
 CC polypeptide (I) that comprises, mimics or cross-reacts with a B- or  
 CC T-cell epitope of a FlgE (flagellar hook) polypeptide from a  
 CC Lawsonia spp. (I) has antibacterial activity, and induces a specific  
 CC humoral immune response. (I) are used as antigens in vaccines to prevent  
 CC or treat infection by Lawsonia, in birds and animals, especially pigs,  
 CC to raise specific antibodies (Ab) and to detect past or present  
 CC infection. Ab are also useful in diagnosis, to detect L. intracellularis  
 CC or immunologically cross-reactive species, also for identification of  
 CC epitopes in FlgE. Vectors that contain nucleic acids (II) encoding (I)  
 CC are also useful in genetic vaccines, and fragments of (II) are useful  
 CC as primers or probes for detecting L. intracellularis or related  
 CC microorganisms, in hybridisation or amplification assays.  
 XX  
 SQ Sequence 486 AA;

Query Match 38.7%; Score 46; DB 22; Length 486;  
 Best Local Similarity 41.2%; Pred. No. 73;  
 Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 1 ANNHVSIVEAWSNDNTP 17  
 I I I I I I I I I I  
 Db 169 anpyfallewkgngtp 185

RESULT 8

AAG60587  
 ID AAG60587 standard; Protein: 174 AA.

AC AAG60587;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 78492.

DE Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.

XX Arabidopsis thaliana.

XX EPI033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127452.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144352.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.

```
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162143.

Query Match 37.8% Score 45; DB 21; Length 174;
Best Local Similarity 50.0%; Pred. No. 31;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 5 VSIVEANSNDTPY 18
Db 122 vvtswsgsstpy 135

RESULT 9
AAG60586
ID AAG60586 standard; Protein; 185 AA.
XX AC AAG60586;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 78491.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX O Arabidopsis thaliana.
XX P EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 23-APR-1999; 99US-0130891.
XX PR 28-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 30-APR-1999; 99US-0132407.
XX PR 04-MAY-1999; 99US-0132484.
XX PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
```



PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140033.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149802.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.





PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144844.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159233.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 37.8%; Score 45; DB 21; Length 305;  
Best Local Similarity 47.1%; Pred.No. 60;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
QY 3 NHVSIVEAWSNDTPYL 19

Db           :: |:|||| | |:  
200 dyeqimeawskgtlyv 216

RESULT 13  
AAG09844  
ID   AAG09844 standard; Protein; 339 AA.  
XX   AAG09844;  
AC   AAG09844;  
XX   17-OCT-2000 (first entry)  
DT   Arabidopsis thaliana protein fragment SEQ ID NO: 7934.  
XX   Arabidopsis thaliana.  
XX   Protein identification; signal transduction pathway; metabolic pathway;  
KW   hybridisation assay; genetic mapping; gene expression control; promoter;  
KW   termination sequence.  
XX   Arabidopsis thaliana.  
OS   EPI033405-A2.  
XX   06-SEP-2000.  
PF   25-FEB-2000; 2000EP-0301439.  
XX   25-FEB-1999; 99US-0121825.  
XX   05-MAR-1999; 99US-0123180.  
XX   09-MAR-1999; 99US-0123548.  
XX   23-MAR-1999; 99US-0125788.  
XX   25-MAR-1999; 99US-0126264.  
XX   29-MAR-1999; 99US-0126785.  
XX   01-APR-1999; 99US-0127462.  
XX   06-APR-1999; 99US-0128234.  
XX   08-APR-1999; 99US-0128714.  
XX   16-APR-1999; 99US-0129845.  
XX   19-APR-1999; 99US-0130077.  
XX   21-APR-1999; 99US-0130449.  
XX   23-APR-1999; 99US-0130510.  
XX   28-APR-1999; 99US-0130891.  
XX   30-APR-1999; 99US-0131449.  
XX   30-APR-1999; 99US-0132048.  
XX   04-MAY-1999; 99US-0132407.  
XX   05-MAY-1999; 99US-0132485.  
XX   06-MAY-1999; 99US-0132486.  
XX   06-MAY-1999; 99US-0132487.  
XX   07-MAY-1999; 99US-0132863.  
XX   11-MAY-1999; 99US-0134256.  
XX   14-MAY-1999; 99US-0134218.  
XX   14-MAY-1999; 99US-0134219.  
XX   14-MAY-1999; 99US-0134221.  
XX   14-MAY-1999; 99US-0134370.  
XX   18-MAY-1999; 99US-0134768.  
XX   19-MAY-1999; 99US-0134941.  
XX   20-MAY-1999; 99US-0135124.  
XX   21-MAY-1999; 99US-0135353.  
XX   24-MAY-1999; 99US-0135629.  
XX   25-MAY-1999; 99US-0136021.  
XX   27-MAY-1999; 99US-0136392.  
XX   28-MAY-1999; 99US-0136782.  
XX   01-JUN-1999; 99US-0137222.  
XX   03-JUN-1999; 99US-0137528.  
XX   04-JUN-1999; 99US-0137502.  
XX   07-JUN-1999; 99US-0137724.  
XX   08-JUN-1999; 99US-0138094.  
XX   10-JUN-1999; 99US-0138540.  
XX   10-JUN-1999; 99US-0138847.  
XX   14-JUN-1999; 99US-0139119.  
XX   16-JUN-1999; 99US-0139452.  
XX   16-JUN-1999; 99US-0139453.  
XX   17-JUN-1999; 99US-0139492.  
XX   18-JUN-1999; 99US-0139454.  
PR   18-JUN-1999; 99US-0139455.  
PR   18-JUN-1999; 99US-0139456.  
PR   18-JUN-1999; 99US-0139457.  
PR   18-JUN-1999; 99US-0139458.  
PR   18-JUN-1999; 99US-0139459.  
PR   18-JUN-1999; 99US-0139460.  
PR   18-JUN-1999; 99US-0139461.  
PR   18-JUN-1999; 99US-0139462.  
PR   18-JUN-1999; 99US-0139463.  
PR   18-JUN-1999; 99US-0139750.  
PR   18-JUN-1999; 99US-0139763.  
PR   21-JUN-1999; 99US-0139817.  
PR   22-JUN-1999; 99US-0139899.  
PR   23-JUN-1999; 99US-0140353.  
PR   23-JUN-1999; 99US-0140354.  
PR   24-JUN-1999; 99US-0140695.  
PR   28-JUN-1999; 99US-0140823.  
PR   29-JUN-1999; 99US-0140991.  
PR   30-JUN-1999; 99US-0141287.  
PR   01-JUL-1999; 99US-0141842.  
PR   01-JUL-1999; 99US-0142154.  
PR   02-JUL-1999; 99US-0142055.  
PR   06-JUL-1999; 99US-0142390.  
PR   08-JUL-1999; 99US-0142803.  
PR   09-JUL-1999; 99US-0142920.  
PR   12-JUL-1999; 99US-0142977.  
PR   13-JUL-1999; 99US-0143542.  
PR   14-JUL-1999; 99US-0143624.  
PR   15-JUL-1999; 99US-0144005.  
PR   16-JUL-1999; 99US-0144085.  
PR   16-JUL-1999; 99US-0144086.  
PR   19-JUL-1999; 99US-0144325.  
PR   19-JUL-1999; 99US-0144331.  
PR   19-JUL-1999; 99US-0144332.  
PR   19-JUL-1999; 99US-0144333.  
PR   19-JUL-1999; 99US-0144334.  
PR   19-JUL-1999; 99US-0144335.  
PR   20-JUL-1999; 99US-0144352.  
PR   20-JUL-1999; 99US-0144632.  
PR   20-JUL-1999; 99US-0144884.  
PR   21-JUL-1999; 99US-0144814.  
PR   21-JUL-1999; 99US-0145086.  
PR   21-JUL-1999; 99US-0145088.  
PR   22-JUL-1999; 99US-0145085.  
PR   22-JUL-1999; 99US-0145087.  
PR   22-JUL-1999; 99US-0145089.  
PR   22-JUL-1999; 99US-0145192.  
PR   23-JUL-1999; 99US-0145145.  
PR   23-JUL-1999; 99US-0145218.  
PR   23-JUL-1999; 99US-0145224.  
PR   26-JUL-1999; 99US-0145276.  
PR   27-JUL-1999; 99US-0145913.  
PR   27-JUL-1999; 99US-0145918.  
PR   27-JUL-1999; 99US-0145919.  
PR   28-JUL-1999; 99US-0145951.  
PR   02-AUG-1999; 99US-0146386.  
PR   02-AUG-1999; 99US-0146388.  
PR   02-AUG-1999; 99US-0146389.  
PR   03-AUG-1999; 99US-0147038.  
PR   04-AUG-1999; 99US-0147204.  
PR   04-AUG-1999; 99US-0147302.  
PR   05-AUG-1999; 99US-0147192.  
PR   05-AUG-1999; 99US-0147260.  
PR   06-AUG-1999; 99US-0147303.  
PR   06-AUG-1999; 99US-0147416.  
PR   09-AUG-1999; 99US-0147493.  
PR   09-AUG-1999; 99US-0147935.  
PR   10-AUG-1999; 99US-0148171.  
PR   11-AUG-1999; 99US-0148319.  
PR   12-AUG-1999; 99US-0148341.  
PR   13-AUG-1999; 99US-0148565.  
PR   13-AUG-1999; 99US-0148684.  
PR   16-AUG-1999; 99US-0149368.



Search completed: March 27, 2002, 13:57:57  
Job time: 525 sec

AC AAW17888;  
XX  
XX 29-JAN-1998 (first entry)  
XX  
XX Photorhabdus luminescens insect toxin protein TcaAii+TcaAiii.  
XX  
XX Insecticide; insect; toxin; pest control; biological control;  
KW Photorhabdus luminescens; TcaA; Southern corn rootworm;  
KW Colorado potato beetle; Western corn rootworm; meal worm;  
KW boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;  
KW cabbage looper; codling moth; corn earworm; European corn borer;  
KW tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;  
KW Diptera, Dictyoptera; Acarina; Homoptera.  
XX  
XX Photorhabdus luminescens strain W-14 (ATCC 55397).  
XX  
XX Key Location/Qualifiers  
FH Protein 254..492  
FT /label= TcaAii  
FT Protein 493..1095  
FT /label= TcaAiii  
XX  
XX W09717432-A1.  
XX  
XX 15-MAY-1997.  
XX  
XX 06-NOV-1996; 96WO-US18003.  
XX  
XX 28-AUG-1996; 96US-0705484.  
XX 06-NOV-1995; 95US-0007255.  
XX 28-FEB-1996; 96US-0608423.  
XX  
XX (WISC ) WISCONSIN ALUMNI RES FOUND.  
XX  
XX Blackburn MB, Bowen DJ, Ciche TA, Ensign JC, Fatig R;  
PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;  
PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;  
PI Strickland JA;  
XX  
XX WPI: 1997-281022/25.  
DR N-PSDB: AAT68841.  
XX  
XX Photorhabdus sp. insecticidal protein toxins and DNA encoding them -  
PT can be genetically engineered into insect larvae food and plants for  
PT insect control  
XX  
XX Claim 34; Page 169-173; 276pp; English.  
XX  
XX This polypeptide comprises a claimed insecticidal toxin protein,  
CC TcaAii+TcaAiii (see also AAW17889), of Photorhabdus luminescens that  
CC is a component of a toxin protein complex. Claimed toxins of P.  
CC luminescens (see AAW17871, AAW17884-89, AAW17899-900, AAW18301-06) can  
CC be applied to, or genetically engineered into, insect larvae food and  
CC plants for insect control. The Photorhabdus toxins are particularly  
CC effective against Southern corn rootworm, Colorado potato beetle,  
CC western corn rootworm, meal worm, boll weevil and turf grub  
CC (Coleoptera), beet armyworm, black cutworm, cabbage looper, codling  
CC moth, corn earworm, European corn borer, tobacco hornworm and  
CC tobacco budworm (Lepidoptera), and are also active against insects  
CC of the orders Hymenoptera, Diptera, Dictyoptera, Acarina and  
CC Homoptera. (All claimed).  
XX  
XX Sequence 1095 AA;  
XX  
XX Query Match 37.8%; Score 45; DB 18; Length 1095;  
Best Local Similarity 33.3%; Pred. No. 2.7e+02;  
Matches 7; Conservative 5; Mismatches 9; Indels 0; Gaps 0;  
QY 1 ANNHVSIVEAWSNDNTPYLHD 21  
Db 887 sdnpflianywhdnqtlfshd 907



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:59:30 ; Search time 87.3 Seconds  
(without alignments)  
5.413 Million cell updates/sec

Title: US-09-290-049A-13  
Perfect score: 113  
Sequence: 1 AIDHLSILEAWSGNDNDYVKQ 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PTCUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	75	66.4	1430	3	US-09-008-172-2	Sequence 2, Appli
2	75	66.4	1430	4	US-09-210-361-6	Sequence 6, Appli
3	73	64.6	1475	3	US-09-007-999-2	Sequence 2, Appli
4	73	64.6	1475	4	US-09-210-361-2	Sequence 2, Appli
5	73	64.6	1577	2	US-08-793-824-2	Sequence 2, Appli
6	71	62.8	1375	4	US-09-210-361-4	Sequence 4, Appli
7	45	39.8	2366	1	US-08-480-604A-10	Sequence 10, Appl
8	45	39.8	2366	2	US-08-405-496A-10	Sequence 10, Appl
9	45	39.8	2366	4	US-08-915-136-10	Sequence 10, Appl
10	44	38.9	270	2	US-08-852-743-5	Sequence 5, Appli
11	44	38.9	270	3	US-09-185-370-5	Sequence 5, Appli
12	44	38.9	487	2	US-08-712-709-8	Sequence 8, Appli
13	44	38.9	487	3	US-09-111-444-8	Sequence 8, Appli
14	44	38.9	487	4	US-09-541-228-8	Sequence 8, Appli
15	44	38.9	639	2	US-08-557-309B-37	Sequence 37, Appl
16	44	38.9	639	3	US-08-834-306-37	Sequence 37, Appl
17	44	38.9	639	4	US-08-993-674A-37	Sequence 37, Appl
18	43	38.1	456	4	US-09-172-841-51	Sequence 51, Appl
19	42	37.2	1205	4	US-09-330-330-1	Sequence 1, Appli
20	41	36.3	132	4	US-09-073-297-16	Sequence 16, Appl
21	40	35.4	43	1	US-08-149-839B-8	Sequence 8, Appli
22	40	35.4	43	1	US-08-451-568-8	Sequence 8, Appli
23	40	35.4	43	1	US-08-451-566-8	Sequence 8, Appli
24	40	35.4	43	2	US-08-777-113-8	Sequence 8, Appli
25	40	35.4	71	1	US-07-704-288C-15	Sequence 15, Appl
26	40	35.4	71	1	US-08-379-259-15	Sequence 15, Appl
27	40	35.4	148	3	US-08-329-799-35	Sequence 35, Appl

28	40	35.4	310	1	US-07-704-288C-6	Sequence 6, Appli
29	40	35.4	310	1	US-08-379-259-6	Sequence 6, Appli
30	40	35.4	310	4	US-07-791-931-6	Sequence 6, Appli
31	40	35.4	314	1	US-07-704-288C-7	Sequence 7, Appli
32	40	35.4	314	1	US-08-379-259-7	Sequence 7, Appli
33	40	35.4	324	1	US-08-047-413-11	Sequence 11, Appl
34	40	35.4	324	3	US-08-229-050-11	Sequence 11, Appl
35	40	35.4	324	3	US-08-801-563-11	Sequence 11, Appl
36	40	35.4	328	4	US-07-791-931-5	Sequence 5, Appli
37	40	35.4	330	1	US-07-704-288C-8	Sequence 8, Appli
38	40	35.4	330	1	US-08-379-259-8	Sequence 8, Appli
39	39	34.5	387	2	US-08-759-581B-4	Sequence 4, Appli
40	39	34.5	464	2	US-08-759-581B-22	Sequence 22, Appl
41	39	34.5	543	4	US-08-697-610-2	Sequence 2, Appli
42	39	34.5	543	4	US-08-349-357-2	Sequence 2, Appli
43	39	34.5	730	2	US-08-696-944-2	Sequence 2, Appli
44	39	34.5	804	4	US-08-981-446B-3	Sequence 3, Appli
45	39	34.5	1115	3	US-08-323-477-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-09-008-172-2  
; Sequence 2, Application US/09008172  
; Patent No. 6127602  
; GENERAL INFORMATION:  
; APPLICANT: NICHOLS, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; FILE OF INVENTION: Latexes in Paper Manufacture  
; FILE REFERENCE: 0358D  
; CURRENT APPLICATION NUMBER: US/09/008,172  
; CURRENT FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-008-172-2

Query Match 66.4%; Score 75; DB 3; Length 1430;  
Best Local Similarity 75.0%; Pred. No. 0.001;  
Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Oy 1 AIDHLSILEAWSGNDNDYVK 20  
Db 495 AINHLISLEAWSNDPQYK 514  
||:||||||| | | | | |

RESULT 2  
US-09-210-361-6  
; Sequence 6, Application US/09210361  
; Patent No. 6284479  
; GENERAL INFORMATION:  
; APPLICANT: NICHOLS, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starches and  
; FILE OF INVENTION: Latexes in Paper Manufacture  
; FILE REFERENCE: 0357CR  
; CURRENT APPLICATION NUMBER: US/09/210,361  
; CURRENT FILING DATE: 1998-12-11  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/009,620  
; EARLIER FILING DATE: 1998-01-20  
; EARLIER APPLICATION NUMBER: 08/485,243  
; EARLIER FILING DATE: 1995-06-07

; EARLIER APPLICATION NUMBER: 09/008,172  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 6  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-210-361-6

Query Match 66.4%; Score 75; DB 4; Length 1430;  
Best Local Similarity 75.0%; Pred. No. 0.001; 4; Indels 0; Gaps 0;  
Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

1 AIDHLSILEAWSGNDNDYK 20  
||| ||||| ||| |||  
495 AINHLISILEAWSNDPQYNK 514

RESULT 3  
US-09-007-999-2  
; Sequence 2, Application US/09007999  
; Patent No. 6087559  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; FILE OF INVENTION: Latexes in Paper Manufacture  
; FILE REFERENCE: 0356D  
; CURRENT APPLICATION NUMBER: US/09/007,999  
; CURRENT FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-007-999-2

Query Match 64.6%; Score 73; DB 3; Length 1475;  
Best Local Similarity 73.7%; Pred. No. 0.0022; 4; Indels 0; Gaps 0;  
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

1 AIDHLSILEAWSGNDNDYK 19  
| ||||| ||| |||  
Db 481 ANDHLSILEAWSNDTPYL 499

RESULT 4  
US-09-210-361-2  
; Sequence 2, Application US/09210361  
; Patent No. 6284479  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starches and  
; FILE OF INVENTION: Latexes in Paper Manufacture  
; FILE REFERENCE: 0357CR  
; CURRENT APPLICATION NUMBER: US/09/210,361  
; CURRENT FILING DATE: 1998-12-11  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/009,620  
; EARLIER FILING DATE: 1998-01-20  
; EARLIER APPLICATION NUMBER: 08/485,243  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/008,172

; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-210-361-2

Query Match 64.6%; Score 73; DB 4; Length 1475;  
Best Local Similarity 73.7%; Pred. No. 0.0022; 4; Indels 0; Gaps 0;  
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AIDHLSILEAWSGNDNDYK 19  
| ||||| ||| |||  
Db 481 ANDHLSILEAWSNDTPYL 499

RESULT 5  
US-08-793-824-2  
; Sequence 2, Application US/08793824  
; Patent No. 5981838  
; GENERAL INFORMATION:  
; APPLICANT: Simpson, Christine Lynn  
; APPLICANT: Giffard, Phillip Morrison  
; APPLICANT: Jacques, Nicholas Anthony  
; TITLE OF INVENTION: Genetic Manipulation of Plants to  
; TITLE OF INVENTION: Increase Stored Carbohydrates  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Griffith Hack & Co  
; STREET: Level 8, 168 Walker Street  
; CITY: No. 5981838th Sydney  
; STATE: New South Wales  
; COUNTRY: Australia  
; ZIP: 2060  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/793,824  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: AU PM7643  
; FILING DATE: 24-AUG-1994  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 61 2 9957 5944  
; TELEFAX: 61 2 957 6288  
; TELEX: 26547  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1577 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Streptococcus salivarius  
US-08-793-824-2

Query Match 64.6%; Score 73; DB 2; Length 1577;  
Best Local Similarity 75.0%; Pred. No. 0.0024; 4; Indels 0; Gaps 0;  
Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AIDHLSILEAWSGNDNDYK 20  
||| ||||| ||| |||



Db 588 AIAHLSILEAWSYNDHOYNK 607

## RESULT 6

US-09-210-361-4  
; Sequence 4, Application US/09210361  
; Patent No. 6284479  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starches and  
; TITLE OF INVENTION: Latexes in Paper Manufacture  
; FILE REFERENCE: 0357CR  
; CURRENT APPLICATION NUMBER: US/09/210,361  
; CURRENT FILING DATE: 1998-12-11  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/009,620  
; EARLIER FILING DATE: 1998-01-20  
; EARLIER APPLICATION NUMBER: 08/485,243  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/008,172  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 1375  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-210-361-4

Query Match 62.8%; Score 71; DB 4; Length 1375;  
Best Local Similarity 73.7%; Pred. No. 0.0043;  
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 AIDHLSILEAWSGNDNDYV 19

Db 507 ANDHLSILEAWSYNDTPYL 525

## RESULT 7

US-08-480-604A-10  
; Sequence 10, Application US/08480604A  
; Patent No. 5736139  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHVE, NISHA V.  
; APPLICANT: FIRCA, JOSEPH R.  
; APPLICANT: STAFFORD, DOUGLAS C.  
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,604A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/422,711  
; FILING DATE: 14-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/405,496  
; FILING DATE: 16-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: INGOLIA, DIANE E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPD-01763  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2366 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-480-604A-10

Query Match 39.8%; Score 45; DB 1; Length 2366;  
Best Local Similarity 44.4%; Pred. No. 1.3e+02;  
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 IDHLSILEAWSGNDNDYV 19

Db 1733 INDLSIRYVWSNDGNDFI 1750

## RESULT 8

US-08-405-496A-10  
; Sequence 10, Application US/08405496A  
; Patent No. 5919655  
; GENERAL INFORMATION:  
; APPLICANT: WILLIAMS, JAMES A.  
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM  
; TITLE OF INVENTION: NEUROTOXIN  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/405,496A  
; FILING DATE: 16-MAR-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,907

```

? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/480,604
? FILING DATE:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/405,496
? FILING DATE: 16-MAR-1995
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/329,154
? FILING DATE: 25-OCT-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/161,907
? FILING DATE: 02-DEC-1993

```

COUNTRY: US  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/852,743  
FILING DATE: 7-MAY-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/016,774  
FILING DATE: 7-MAY-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Fraser, Janis K.  
REGISTRATION NUMBER: 34,819  
REFERENCE/DOCKET NUMBER: 00786/327001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 270 amino acids  
TYPE: amino acid

; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-852-743-5

Query Match 38.9%; Score 44; DB 2; Length 270;  
Best Local Similarity 61.5%; Pred. No. 14;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 9 EAWSGNDNDYVKQ 21  
| | | | |  
Db 223 ELWSDNFTDFVKQ 235

## RESULT 11

US-09-185-370-5  
; Sequence 5, Application US/09185370  
; Patent No. 6093560  
; GENERAL INFORMATION:  
; APPLICANT: Force, Thomas  
; APPLICANT: Kyriakis, John M.  
; APPLICANT: Pombo, Celia M.  
; APPLICANT: Bonventre, Joseph  
; TITLE OF INVENTION: SOK-1 AND METHODS OF USE  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson, P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/185,370  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/852,743  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fraser, Janis K.  
; REGISTRATION NUMBER: 34,819  
; REFERENCE/DOCKET NUMBER: 00786/327001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 270 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-185-370-5

Query Match 38.9%; Score 44; DB 3; Length 270;  
Best Local Similarity 61.5%; Pred. No. 14;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 9 EAWSGNDNDYVKQ 21  
| | | | |  
Db 223 ELWSDNFTDFVKQ 235

## RESULT 12

US-08-712-709-8  
; Sequence 8, Application US/08712709

; Patent No. 5863780  
; GENERAL INFORMATION:  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Hawkins, Phillip R.  
; TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASES  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: U.S.  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/712,709  
; FILING DATE: Filed Herewith  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0118 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 487 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 1117791  
US-08-712-709-8

Query Match 38.9%; Score 44; DB 2; Length 487;  
Best Local Similarity 61.5%; Pred. No. 28;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 9 EAWSGNDNDYVKQ 21  
| | | | |  
Db 248 ELWSDNFTDFVKQ 260

## RESULT 13

US-09-111-444-8  
; Sequence 8, Application US/09111444  
; Patent No. 6045792  
; GENERAL INFORMATION:  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Hawkins, Phillip R.  
; TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASES  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: U.S.  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ Version 1.5

; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/111.444  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/712.709  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0118 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 487 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 1117791  
US-09-111-444-8

Query Match 38.9%; Score 44; DB 3; Length 487;  
Best Local Similarity 61.5%; Pred. No. 28;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Oy 9 EAWSGNDNDYVKQ 21  
| | | | |  
Db 248 ELWSDNFTDFVKQ 260

RESULT 14  
US-09-541-228-8  
; Sequence 8, Application US/09541228  
; Patent No. 6232077  
; GENERAL INFORMATION:  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Hawkins, Phillip R.  
; TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASES  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: U.S.  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/541.228  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/712.709  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0118 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 487 amino acids

; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 1117791  
US-09-541-228-8

Query Match 38.9%; Score 44; DB 4; Length 487;  
Best Local Similarity 61.5%; Pred. No. 28;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Oy 9 EAWSGNDNDYVKQ 21  
| | | | |  
Db 248 ELWSDNFTDFVKQ 260

RESULT 15  
US-08-557-309B-37  
; Sequence 37, Application US/08557309B  
; Patent No. 5916572  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF  
; NUMBER OF SEQUENCES: 69  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/557.309B  
; FILING DATE: 14-NOV-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.422  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 37:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 639 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-08-557-309B-37

Query Match 38.9%; Score 44; DB 2; Length 639;  
Best Local Similarity 35.3%; Pred. No. 39;  
Matches 6; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

Oy 4 HLSILEAWSGNDNDYVK 20  
| | | | |  
Db 379 NVKLVDAYRGNGNEYVR 395

Search completed: March 27, 2002, 13:59:31

Job time: 584 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: March 27, 2002, 14:20:26 ; Search time 1139.61 seconds  
(without alignments)  
5.116 Million cell updates/sec

Title: US-09-290-049a-13

Perfect score: 113

Sequence: 1 AIDHLSILEAWSGNDNDYVKQ 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 3148936 seqs, 277657034 residues

Total number of hits satisfying chosen parameters: 3148936

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

Pending\_Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/2/paa/PCTUS\_COMB.pep.\*

2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep.\*

3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep.\*

4: /cgn2\_6/ptodata/2/paa/US08\_COMB.pep.\*

5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pep.\*

6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pep.\*

7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pep.\*

8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep.\*

9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pep.\*

10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep.\*

11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pep.\*

12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pep.\*

13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep.\*

14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep.\*

15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pep.\*

16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pep.\*

17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pep.\*

18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pep.\*

19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pep.\*

20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pep.\*

21: /cgn2\_6/ptodata/2/paa/US097\_COMB.pep.\*

22: /cgn2\_6/ptodata/2/paa/US098\_COMB.pep.\*

23: /cgn2\_6/ptodata/2/paa/US099\_COMB.pep.\*

24: /cgn2\_6/ptodata/2/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	113	100.0	21	16	US-09-290-049-13
2	75	66.4	21	16	US-09-290-049-11
3	75	66.4	1430	20	US-09-649-885-2
4	75	66.4	1430	21	US-09-740-274-6
5	73	64.6	21	16	US-09-290-049-1
6	73	64.6	21	16	US-09-290-049-10
7	73	64.6	1475	19	US-09-557-848-2
8	73	64.6	1475	21	US-09-740-274-2
9	71	62.8	1375	21	US-09-740-274-4

10	65	57.5	2057	18	US-09-499-203-2	Sequence 2, Appl
11	62	54.9	21	16	US-09-290-049-12	Sequence 12, Appl
12	62	54.9	21	16	US-09-290-049-14	Sequence 14, Appl
13	49.5	43.8	528	18	US-09-489-039A-11958	Sequence 11958, A
14	49	43.4	838	18	US-09-401-978C-66	Sequence 66, Appl
15	48.5	42.9	289	1	PCT-US00-35017A-1386	Sequence 1386, Appl
16	48.5	42.9	319	1	PCT-US01-08656-9177	Sequence 9177, Ap
17	48.5	42.9	426	20	US-09-628-508-98	Sequence 98, Appl
18	48.5	42.9	427	1	PCT-US00-03062-98	Sequence 98, Appl
19	48.5	42.9	427	1	PCT-US00-03062-98	Sequence 98, Appl
20	48.5	42.9	427	1	PCT-US99-01621-129	Sequence 129, App
21	48.5	42.9	427	17	US-09-363-044A-131	Sequence 131, App
22	48.5	42.9	427	23	US-09-949-925-131	Sequence 131, App
23	48.5	42.9	435	20	US-09-628-508-97	Sequence 97, Appl
24	48.5	42.9	436	1	PCT-US00-03062-97	Sequence 97, Appl
25	48.5	42.9	436	1	PCT-US00-03062-97	Sequence 97, Appl
26	48.5	42.9	576	1	PCT-US00-03062-70	Sequence 70, Appl
27	48.5	42.9	576	1	PCT-US00-03062-70	Sequence 70, Appl
28	48.5	42.9	576	20	US-09-628-508-70	Sequence 70, Appl
29	48	42.5	366	21	US-09-739-449-9220	Sequence 9220, Ap
30	48	42.5	366	22	US-09-803-110-9220	Sequence 9220, Ap
31	48	42.5	661	15	US-09-198-452A-36	Sequence 36, Appl
32	48	42.5	705	18	US-09-438-185-22	Sequence 22, Appl
33	47.5	42.0	310	16	US-09-248-796-17322	Sequence 17322, A
34	45	39.8	58	1	PCT-US01-01324-1280	Sequence 1280, Ap
35	45	39.8	58	1	PCT-US01-01351-163	Sequence 163, App
36	45	39.8	58	21	US-09-764-887-163	Sequence 163, App
37	45	39.8	226	18	US-09-417-507-41977	Sequence 41977, A
38	45	39.8	409	20	US-09-619-049-765	Sequence 765, App
39	45	39.8	409	24	US-60-167-217-15139	Sequence 15139, A
40	45	39.8	409	24	US-60-171-627-1224	Sequence 1224, Ap
41	45	39.8	409	24	US-60-173-464-12341	Sequence 12341, A
42	45	39.8	409	24	US-60-191-637-15101	Sequence 15101, A
43	45	39.8	409	24	US-60-191-681-11930	Sequence 11930, A
44	45	39.8	412	1	PCT-US01-08631-44060	Sequence 44060, A
45	45	39.8	440	1	PCT-US01-08631-53472	Sequence 53472, A

## ALIGNMENTS

RESULT 1

US-09-290-049-13

Sequence 13, Application US/09290049

GENERAL INFORMATION:

APPLICANT: Smith, Daniel J.

APPLICANT: Taubman, Martin A.

TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL

TITLE OF INVENTION: CARRIES

FILE REFERENCE: PDC98-01p2A

CURRENT APPLICATION NUMBER: US/09/290,049

CURRENT FILING DATE: 1999-04-12

EARLIER APPLICATION NUMBER: 60/081,550

EARLIER FILING DATE: 1998-04-13

EARLIER APPLICATION NUMBER: 60/115,142

EARLIER FILING DATE: 1999-01-08

NUMBER OF SEQ ID NOS: 19

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 13

LENGTH: 21

TYPE: PRT

ORGANISM: S. downei

US-09-290-049-13

Query Match 100.0%; Score 113; DB 16; Length 21;

Best Local Similarity 100.0%; Pred. No. 3e-10;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AIDHLSILEAWSGNDNDYVKQ 21

Db 1 AIDHLSILEAWSGNDNDYVKQ 21

```
RESULT 2
US-09-290-049-11
; Sequence 11, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CRIES
; FILE REFERENCE: FDC98-01p2A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. mutans
US-09-290-049-11
```

```
Query Match 66.4%; Score 75; DB 16; Length 21;
Best Local Similarity 75.0%; Pred. No. 0.00027;
Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 1 AIDHLSILEAWSGNDNDYVK 20
||:|||||  ||
Db 1 AINHLSILEAWSNDNDPQYNK 20
```

```
RESULT 3
US-09-649-885-2
; Sequence 2, Application US/09649885
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0358D2
; CURRENT APPLICATION NUMBER: US/09/649,885
; CURRENT FILING DATE: 2000-08-28
; PRIOR APPLICATION NUMBER: US 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-649-885-2
```

```
Query Match 66.4%; Score 75; DB 20; Length 1430;
Best Local Similarity 75.0%; Pred. No. 0.026;
Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 1 AIDHLSILEAWSGNDNDYVK 20
||:|||||  ||
Db 495 AINHLSILEAWSNDNDPQYNK 514
```

```
RESULT 4
US-09-740-274-6
; Sequence 6, Application US/09740274
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
```

```
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-6
```

```
Query Match 66.4%; Score 75; DB 21; Length 1430;
Best Local Similarity 75.0%; Pred. No. 0.026;
Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 1 AIDHLSILEAWSGNDNDYVK 20
||:|||||  ||
Db 495 AINHLSILEAWSNDNDPQYNK 514
```

```
RESULT 5
US-09-290-049-1
; Sequence 1, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CRIES
; FILE REFERENCE: FDC98-01p2A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: EAW peptide
US-09-290-049-1
```

```
Query Match 64.6%; Score 73; DB 16; Length 21;
Best Local Similarity 73.7%; Pred. No. 0.00055;
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 1 AIDHLSILEAWSGNDNDYV 19
|:|||||  |
Db 1 AINHLSILEAWSNDNDTPYL 19
```

```
RESULT 6
US-09-290-049-10
; Sequence 10, Application US/09290049
; GENERAL INFORMATION:
```



```

RESULT      8
US-09-740-274-2
; Sequence 2, Application US/09740274
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999

```

```

RESULT          9
US-09-740-274-4
; Sequence 4, Application US/09740274
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-4

```

Qy 1 AIDHLSILEAWSGNNDYV 19  
| | | | | | | | | | :  
Db 507 ANDHLSILEAWSYNDDPYL 525

RESULT 10  
US-09-499-203-2  
; Sequence 2; Application US/09499203



Best Local Similarity 50.0%; Pred. No. 1.7e+02;  
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;  
Qy 4 HLSILEAWSGNDNDYV 19  
Db 623 HLTIFHSLQGNKDYI 638

RESULT 15  
PCT-US00-35017A-1386  
; Sequence 1386, Application PC/TUS0035017A  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq Inc  
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides  
; FILE REFERENCE: 784PCT  
; CURRENT APPLICATION NUMBER: PCT/US00/35017A  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; PRIOR APPLICATION NUMBER: US09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; NUMBER OF SEQ ID NOS: 1478  
; SEQ ID NO 1386  
; LENGTH: 289  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
PCT-US00-35017A-1386

Query Match 42.9%; Score 48.5; DB 1; Length 289;  
Best Local Similarity 52.6%; Pred. No. 65;  
Matches 10; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

Qy 3 DHLSTLEAWSGNDNDYVKQ 21  
Db 269 DRLVTLESWA-NDDPYLKR 286

Search completed: March 27, 2002, 14:20:27  
Job time: 1574 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:22:47 ; Search time 137.48 seconds  
(without alignments)  
10.540 Million cell updates/sec

Title: US-09-290-049A-13  
Perfect score: 113  
Sequence: 1 AIDHLSILEAWSGNDNDYVKQ 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 332938 seqs, 68999538 residues

Total number of hits satisfying chosen parameters: 332938

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_New.\*

- 1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep.\*
- 7: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep.\*
- 8: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	113	100.0	21	6	US-09-562-328-24
2	113	100.0	21	6	US-09-290-049A-13
3	75	66.4	21	6	US-09-562-328-22
4	75	66.4	21	6	US-09-290-049A-11
5	75	66.4	545	6	US-09-604-957-4
6	73	64.6	21	6	US-09-562-328-20
7	73	64.6	21	6	US-09-290-049A-1
8	73	64.6	21	6	US-09-290-049A-10
9	71	62.8	21	6	US-09-562-328-21
10	65	57.5	584	6	US-09-604-957-6
11	62	54.9	21	6	US-09-562-328-23
12	62	54.9	21	6	US-09-562-328-25
13	62	54.9	21	6	US-09-290-049A-12
14	62	54.9	21	6	US-09-290-049A-14
15	58	51.3	523	6	US-09-604-957-5
16	48.5	42.9	426	6	US-09-997-131-98
17	48.5	42.9	435	6	US-09-997-131-97
18	48.5	42.9	576	6	US-09-997-131-70
19	48	42.5	366	6	US-09-708-427-27416
20	48	42.5	367	6	US-09-708-427-27415
21	48	42.5	378	6	US-09-708-427-27414
22	45	39.8	58	7	US-10-073-961-163
23	45	39.8	58	7	US-10-079-979-1280
24	45	39.8	409	6	US-09-614-150-15057
25	45	39.8	2366	7	US-10-011-366-10

26	44	38.9	297	6	US-09-926-036-6	Sequence 6, Appli
27	44	38.9	357	6	US-09-760-446A-2093	Sequence 2093, Ap
28	44	38.9	388	6	US-09-760-446A-1431	Sequence 1431, Ap
29	44	38.9	453	6	US-09-614-150-25923	Sequence 25923, A
30	43	38.1	535	6	US-09-604-957-7	Sequence 7, Appli
31	43	38.1	766	6	US-09-614-150-15486	Sequence 15486, A
32	43	38.1	1151	1	PCT-US01-25701A-33	Sequence 33, Appl
33	43	38.1	1151	6	US-09-921-159-33	Sequence 33, Appl
34	43	38.1	1235	1	PCT-US01-25701A-2	Sequence 2, Appli
35	43	38.1	1235	6	US-09-921-159-2	Sequence 2, Appli
36	43	38.1	1237	1	PCT-US01-25701A-32	Sequence 32, Appl
37	43	38.1	1237	6	US-09-921-159-32	Sequence 32, Appl
38	43	38.1	1237	7	US-10-024-623-32	Sequence 32, Appl
39	43	38.1	1278	6	US-09-604-957-3	Sequence 3, Appli
40	42.5	37.6	397	6	US-09-708-427-65286	Sequence 65286, A
41	42.5	37.6	458	6	US-09-708-427-65285	Sequence 65285, A
42	42	37.2	116	7	US-10-078-770-196	Sequence 196, App
43	42	37.2	160	6	US-09-860-232-17	Sequence 17, Appl
44	42	37.2	160	6	US-09-860-232-17	Sequence 17, Appl
45	42	37.2	188	6	US-09-675-784A-8550	Sequence 8550, Ap

ALIGNMENTS

RESULT 1

US-09-562-328-24

; Sequence 24, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 24  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-24

Query Match 100.0%; Score 113; DB 6; Length 21;  
Best Local Similarity 100.0%; Pred. No. 7.5e-12;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AIDHLSILEAWSGNDNDYVKQ 21  
Db 1 AIDHLSILEAWSGNDNDYVKQ 21

RESULT 2

US-09-290-049A-13  
; Sequence 13, Application US/09290049A  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: S. downei  
US-09-290-049A-13

Query Match 100.0%; Score 113; DB 6; Length 21;  
Best Local Similarity 100.0%; Pred. No. 7.5e-12;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AIDHLSILEAWSGNDNDYVK 21  
|:||||||||| | | | |  
Db 1 AIDHLSILEAWSGNDNDYVK 21

## RESULT 3

US-09-562-328-22

; Sequence 22, Application US/09562328

; GENERAL INFORMATION:

; APPLICANT: LEES, ANDREW

; APPLICANT: TAUBMAN, MARTIN A.

; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES

; FILE REFERENCE: 04995.0046-01

; CURRENT APPLICATION NUMBER: US/09/562,328

; CURRENT FILING DATE: 2000-05-01

; PRIOR APPLICATION NUMBER: 09/288,965

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 44

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 22

; LENGTH: 21

; TYPE: PRT

; ORGANISM: Streptococcus sp.

US-09-562-328-22

## Query Match

Best Local Similarity 66.4%; Score 75; DB 6; Length 21;

Best Local Similarity 75.0%; Pred. No. 1.1e-05;

Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 AIDHLSILEAWSGNDNDYVK 20  
|:||||||||| | | | |  
Db 1 AINHLISLEAWSNDPOYNK 20

## RESULT 4

US-09-290-049A-11

; Sequence 11, Application US/09290049A

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J.

; APPLICANT: Taubman, Martin A.

; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL

; TITLE OF INVENTION: CARIES

; FILE REFERENCE: 1564.1008-002

; CURRENT APPLICATION NUMBER: US/09/290,049A

; CURRENT FILING DATE: 1999-04-12

; PRIOR APPLICATION NUMBER: 60/081,550

; PRIOR FILING DATE: 1998-04-13

; PRIOR APPLICATION NUMBER: 60/115,142

; PRIOR FILING DATE: 1999-01-08

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 11

; LENGTH: 21

; TYPE: PRT

; ORGANISM: S. mutans

US-09-290-049A-11

## Query Match

Best Local Similarity 66.4%; Score 75; DB 6; Length 21;

Best Local Similarity 75.0%; Pred. No. 1.1e-05;

Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 1 AIDHLSILEAWSGNDNDYVK 20  
|:||||||||| | | | |  
Db 1 AINHLISLEAWSNDPOYNK 20

## RESULT 5

US-09-604-957-4

; Sequence 4, Application US/09604957

; GENERAL INFORMATION:

; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA

; APPLICANT: DIJKHUIZEN, LUBBERT

; APPLICANT: RAHAOUI, HAKIM

; APPLICANT: LEER, ROBERT-JAN

; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN

; FILE REFERENCE: BO 43388

; CURRENT APPLICATION NUMBER: US/09/604,957

; CURRENT FILING DATE: 2000-06-28

; PRIOR APPLICATION NUMBER: 00201871.1

; PRIOR FILING DATE: 2000-05-25

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 4

; LENGTH: 545

; TYPE: PRT

; ORGANISM: Streptococcus mutans

US-09-604-957-4

## Query Match

Best Local Similarity 66.4%; Score 75; DB 6; Length 545;

Best Local Similarity 75.0%; Pred. No. 0.00052;

Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 AIDHLSILEAWSGNDNDYVK 20  
|:||||||||| | | | |  
Db 75 AINHLISLEAWSNDPOYNK 94

## RESULT 6

US-09-562-328-20

; Sequence 20, Application US/09562328

; GENERAL INFORMATION:

; APPLICANT: LEES, ANDREW

; APPLICANT: TAUBMAN, MARTIN A.

; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES

; FILE REFERENCE: 04995.0046-01

; CURRENT APPLICATION NUMBER: US/09/562,328

; CURRENT FILING DATE: 2000-05-01

; PRIOR APPLICATION NUMBER: 09/288,965

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 44

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 20

; LENGTH: 21

; TYPE: PRT

; ORGANISM: Streptococcus sp.

US-09-562-328-20

## Query Match

Best Local Similarity 64.6%; Score 73; DB 6; Length 21;

Best Local Similarity 73.7%; Pred. No. 2.2e-05;

Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 AIDHLSILEAWSGNDNDYV 19  
|:||||||||| | | | |  
Db 1 AINHLISLEAWSNDTPYL 19

## RESULT 7

US-09-290-049A-1

; Sequence 1, Application US/09290049A

; GENERAL INFORMATION:  
; APPLICANT: Taubman, Daniel J.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CARIES  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: EAW peptide  
US-09-290-049A-1

Query Match 64.6%; Score 73; DB 6; Length 21;  
Best Local Similarity 73.7%; Pred. No. 2.2e-05;  
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AIDHLSILEAWSGNDNDYV 19  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 ANDHLSILEAWSNDNDTPYL 19

RESULT 8  
US-09-290-049A-10  
; Sequence 10, Application US/09290049A  
; GENERAL INFORMATION:  
; APPLICANT: Taubman, Daniel J.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CARIES  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; EQ ID NO 10  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: S. mutans  
US-09-290-049A-10

Query Match 64.6%; Score 73; DB 6; Length 21;  
Best Local Similarity 73.7%; Pred. No. 2.2e-05;  
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AIDHLSILEAWSGNDNDYV 19  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 ANDHLSILEAWSNDNDTPYL 19

RESULT 9  
US-09-562-328-21  
; Sequence 21, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES

; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 21  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-21

Query Match 62.8%; Score 71; DB 6; Length 21;  
Best Local Similarity 73.7%; Pred. No. 4.7e-05;  
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AIDHLSILEAWSGNDNDYV 19  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 ANDHLSILEAWSNDNDTPYL 19

RESULT 10  
US-09-604-957-6  
; Sequence 6, Application US/09604957  
; GENERAL INFORMATION:  
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
; APPLICANT: DIJKHUIZEN, LOBBERT  
; APPLICANT: RAHAOUI, HAKIM  
; APPLICANT: LEER, ROBERT-JAN  
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
; FILE REFERENCE: BO 43388  
; CURRENT APPLICATION NUMBER: US/09/604,957  
; CURRENT FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 00201871.1  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 584  
; TYPE: PRT  
; ORGANISM: Leuconostoc mesenteroides  
US-09-604-957-6

Query Match 57.5%; Score 65; DB 6; Length 584;  
Best Local Similarity 66.7%; Pred. No. 0.023;  
Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 4 HLSILEAWSGNDNDYVKQ 21  
| | | | | | | | | | | | | | | | | | | | | |  
Db 78 HLSILEDWNGKDPQYVQ 95

RESULT 11  
US-09-562-328-23  
; Sequence 23, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 23  
; LENGTH: 21  
; TYPE: PRT

; ORGANISM: Streptococcus sp.  
US-09-562-328-23

Query Match 54.9%; Score 62; DB 6; Length 21;  
Best Local Similarity 57.9%; Pred. NO. 0.0014;  
Matches 11; Conservative 4; Mismatches 4; Indels

QY 1 AIDHLSILEAWSGNDNDYV 19  
| : | : | : | : | : | :  
Db 1 ANNHVSIVEAWSNDTPLYL 19

```

RESULT 12
US-09-562-328-25
; Sequence 25, Application US/09562328
; GENERAL INFORMATION:
; APPLICANT: LEES, ANDREW
; APPLICANT: TAUBMAN, MARTIN A.
; APPLICANT: SMITH, DANIEL J.
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES
; FILE REFERENCE: 04995.0046-01
; CURRENT APPLICATION NUMBER: US/09/562,328
; CURRENT FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 09/288,965
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Streptococcus sp.
US-09-562-328-25

```

Query Match 54.9%; Score 62; DB 6; Length 21;  
Best Local Similarity 57.9%; Pred. NO. 0.0014;  
Matches 11; Conservative 4; Mismatches 4; Indels

**Qy**      1 AIDHLSLEAWSGNDNDYV 19  
         | : |:|:|:|:|:|:|:|:

**Dd**      1 ANNHVSIVEAWSDNDTPYL 19

```

RESULT 13
US-09-290-049A-12
Sequence 12, Application US/09290049A
GENERAL INFORMATION:
APPLICANT: Smith, Daniel J.
APPLICANT: Taubman, Martin A.
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
FILE OF INVENTION: CARRIES
FILE REFERENCE: 1564.1008-002
CURRENT APPLICATION NUMBER: US/09/290, 049A
CURRENT FILING DATE: 1999-04-12
PRIOR APPLICATION NUMBER: 60/081,550
PRIOR FILING DATE: 1998-04-13
PRIOR APPLICATION NUMBER: 60/115,142
PRIOR FILING DATE: 1999-01-08
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12
LENGTH: 21
TYPE: PRT
ORGANISM: S. downei
US-09-290-049A-12

```

Query Match 54.9%; Score 62; DB 6; Length 21;  
Best Local Similarity 57.9%; Pred. No. 0.0014;  
Matches 11; Conservative 4; Mismatches 4; Indels

QY 1 AIDHLSILEAWSGNDNDYV 19

db 1 ANNHVSIVEAWSNDTPYL 19

```

RESULT 14
US-09-290-049A-14
; Sequence 14, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: faubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; ; FILE OF INVENTION: CARRIES
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290,049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. sobrinus
US-09-290-049A-14

```

Query Match	54.9%	Score 62;	DB 6;	Length 21;
Best Local Similarity	57.9%	Pred. No. 0.0014;		
Matches 11; Conservative	4;	Mismatches 4;	Indels 0;	Gaps 0;

Qy 1 AIDHLSILEAWSGNDNDYV 19  
| : | : | | | | :  
pb 1 ANNHVSTVEAWSNDNDTPVL 19

```

RESULT 15
US-09-604-957-5
; Sequence 5, Application US/09604957
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHROUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
; US-09-604-957-5

```

Query Match 51.3%; Score 58; DB 6;  
Best Local Similarity 68.4%; Pred. No. 0.28;  
Matches 13: Conservative 0; Mismatches 6; Indels

Qy 1 AIDHLSILEAWSGNDNDYV 19  
| | | | | | | | | |  
Db 75 ANOHLSELDWSHNDPLYV 93

Search completed: March 27, 2002, 14:22:48  
Job time: 1695 sec



---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:01:20 ; Search time 102.51 Seconds  
(without alignments)  
15.605 Million cell updates/sec

Title: US-09-290-049a-13  
Perfect score: 113  
Sequence: 1 AIDHLSILEAWSGNDNDYVK 21  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	108	95.6	1365	2 A41483	glucosyltransferas
2	75	66.4	1431	2 A45866	dextranucrase (EC
3	73	64.6	1475	2 B33135	gtfB protein precu
4	73	64.6	1577	2 T30858	glucosyltransferas
5	71	62.8	1375	2 JT0345	dextranucrase (EC
6	69	61.1	1508	2 T31098	probable dextranu
7	69	61.1	1518	2 A44811	glucosyltransferas
8	67	59.3	1449	2 T30857	glucosyltransferas
9	67	59.3	1449	2 T30552	glucosyltransferas
10	62	54.9	1592	2 A38175	glucosyltransferas
11	61	54.0	1599	2 S22737	glucosyltransferas
12	49	43.4	232	2 B84176	glucosyltransferas
13	49	43.4	1428	1 TSBYT2	hypothetical prote
14	48	42.5	378	2 T04254	DNA topoisomerase
15	48	42.5	695	2 B72129	hypothetical prote
16	48	42.5	695	2 F86493	probable outer mem
17	48	42.5	704	2 C81542	conserved hypothet
18	48	42.5	3206	1 GNV3PV	genome polyprotein
19	47	41.6	298	2 T15906	hypothetical prote
20	47	41.6	680	2 T39858	hypothetical prote
21	47	41.6	769	2 D86335	hypothetical prote
22	46	40.7	287	2 T36413	probable ABC-type
23	46	40.7	458	2 H86314	hypothetical prote
24	46	40.7	536	2 T13261	hypothetical prote
25	46	40.7	1802	2 T00020	bacterial blight-r
26	45	39.8	2366	2 S10317	toxin B - Clostrid
27	45	39.8	2367	2 S70172	toxin B - Clostrid
28	44	38.9	342	2 E69581	acetoin dehydrogen
29	44	38.9	389	2 T17601	hypothetical prote

30 44 38.9 468 2 S61964 probable membrane  
31 44 38.9 588 2 C83836 subtilisin-type pr  
32 43.5 38.5 356 2 T48354 hypothetical prote  
33 43 38.1 76 2 T42309 hypothetical prote  
34 43 38.1 177 2 T40825 hypothetical prote  
35 43 38.1 272 2 H72596 hypothetical prote  
36 43 38.1 429 2 T45040 probable outer mem  
37 43 38.1 697 2 H71525 hypothetical prote  
38 43 38.1 700 2 B81682 conserved hypothet  
39 43 38.1 933 2 T25600 hypothetical prote  
40 43 38.1 1237 2 T46609 calcium-activated  
41 42.5 37.6 69 2 D84933 cof protein (impor  
42 42.5 37.6 137 2 S65216 hypothetical prote  
43 42.5 37.6 140 2 B86683 prophage pil prote  
44 42 37.2 235 2 B70530 hypothetical prote  
45 42 37.2 347 2 T48610 hypothetical prote

ALIGNMENTS

RESULT 1

A41483  
glucosyltransferase (EC 2.4.1.-) gtfS precursor - Streptococcus sobrinus  
C:Species: Streptococcus sobrinus  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 15-Oct-1999  
C:Accession: A41483  
R:Gillmore, K.S.; Russell, R.R.B.; Ferretti, J.J.  
Infect. Immun. 58, 2452-2458, 1990  
A:Title: Analysis of the Streptococcus downei gtfS gene, which specifies a glucosyltr  
A:Reference number: A41483; MUID:90316665  
A:Accession: A41483  
A:Molecule type: DNA  
A:Residues: 1-1365 <GIL>  
A:Cross-references: GB:M30943; NID:gl53652; PIDN:AAA26898.1; PID:gl53653  
C:Gene: gtfS  
C:Superfamily: cpl repeat homology  
C:Keywords: glucosyltransferase; hexosyltransferase

Query Match 95.6%; Score 108; DB 2; Length 1365;  
Best Local Similarity 100.0%; Pred. No. 9.5e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AIDHLSILEAWSGNDNDYVK 20  
Db 467 AIDHLSILEAWSGNDNDYVK 486

RESULT 2

A45866  
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans  
C:Species: Streptococcus mutans  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: A45866  
R:Honda, O.; Kato, C.; Kuramitsu, H.K.  
J. Gen. Microbiol. 136, 2099-2105, 1990  
A:Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the gluco  
A:Reference number: A45866; MUID:91100958  
A:Accession: A45866  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1431 <HON>  
A:Cross-references: GB:M29296  
C:Superfamily: cpl repeat homology  
C:Keywords: glucosyltransferase; hexosyltransferase  
F:181-201/Domain: cpl repeat homology <Cpl>  
F:1127-1146/Domain: cpl repeat homology <CP2>  
F:1192-1211/Domain: cpl repeat homology <CP3>  
F:1257-1276/Domain: cpl repeat homology <CP4>  
F:1277-1297/Domain: cpl repeat homology <CP5>  
F:1321-1340/Domain: cpl repeat homology <CP8>

F:1341-1361/Domain: cpl repeat homology <CP6>  
F:1385-1404/Domain: cpl repeat homology <CP7>

Query Match 66.4%; Score 75; DB 2; Length 1431;  
Best Local Similarity 75.0%; Pred. No. 0.0017;  
Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 AIDHLSILEAWSGNDNDYVK 20  
||:|||||  
Db 495 AINHLISLEAWSDNDPQYNK 514

## RESULT 3

gtfB protein precursor - Streptococcus mutans

C:Species: Streptococcus mutans  
C:Date: 23-Oct-1990 #sequence\_revision 23-Oct-1990 #text\_change 15-Oct-1999

A:Accession: B33135

R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

J. Bacteriol. 169, 4263-4270, 1987

A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.

A:Reference number: A33135; MUID:87308013

A:Accession: B33135

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1475 <SHI>

A:Cross-references: GB:M17361; NID:g153639; PIDN:AAA89588.1; PID:g153640

R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

submitted to The Protein Sequence Database, September 1990

A:Reference number: A33128

A:Accession: A33128

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-171,173-641,'N',643-1475 <SH2>

A:Experimental source: strain GS-5

C:Superfamily: cpl repeat homology

F:1096-1115/Domain: cpl repeat homology <CP1>

F:1224-1243/Domain: cpl repeat homology <CP2>

F:1289-1308/Domain: cpl repeat homology <CP3>

F:1354-1373/Domain: cpl repeat homology <CP4>

F:1419-1438/Domain: cpl repeat homology <CP5>

Query Match 64.6%; Score 73; DB 2; Length 1475;  
Best Local Similarity 73.7%; Pred. No. 0.0037;  
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 AIDHLSILEAWSGNDNDYVK 19  
||:|||||  
Db 481 ANDHLSILEAWSDNDTPYL 499

## RESULT 4

glucosyltransferase - Streptococcus salivarius

C:Species: Streptococcus salivarius

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999

C:Accession: T30858

R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.

Infect. Immun. 63, 609-621, 1995

A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pri

A:Reference number: 220909; MUID:95122197

A:Accession: T30858

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1577 <SIM>

A:Cross-references: EMBL:L35928; NID:g662380; PID:g662381; PIDN:AAC41413.1

C:Genetics:

A:Gene: gtfm

Query Match 64.6%; Score 73; DB 2; Length 1577;

Best Local Similarity 75.0%; Pred. No. 0.004;  
Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 AIDHLSILEAWSGNDNDYVK 20  
||:|||||  
Db 588 AIAHLSILEAWSYNDHQYNK 607

## RESULT 5

dextrantransferase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)

N:Alternate names: sucrose 6-glucosyltransferase

C:Species: Streptococcus mutans

C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 24-Sep-1999

C:Accession: JT0345; C33135

R:Ueda, S.; Shiroza, T.; Kuramitsu, H.K.

Gene 69, 101-109, 1988

A:Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.

A:Reference number: JT0345; MUID:89137980

A:Accession: JT0345

A:Molecule type: DNA

A:Residues: 1-1375 <UED>

A:Experimental source: GS-5

R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

J. Bacteriol. 169, 4263-4270, 1987

A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.

A:Reference number: A33135; MUID:87308013

A:Accession: C33135

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-349 <SHI>

A:Cross-references: GB:M17361

C:Genetics:

A:Gene: gtfC

C:Function:

A:Description: catalyzes the synthesis of both water-soluble and water-insoluble gluc

C:Superfamily: cpl repeat homology

C:Keywords: duplication; glucosyltransferase; hexosyltransferase

F:1-34/Domain: signal sequence #status predicted <SIG>

F:35-1375/Product: glucosyltransferase #status predicted <SIG>

F:1126-1145/Domain: cpl repeat homology <CP1>

F:1253-1272/Domain: cpl repeat homology <CP2>

F:1318-1337/Domain: cpl repeat homology <CP3>

Query Match 62.8%; Score 71; DB 2; Length 1375;  
Best Local Similarity 73.7%; Pred. No. 0.0071;  
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 AIDHLSILEAWSGNDNDYVK 19  
||:|||||  
Db 507 ANDHLSILEAWSYNDTPYL 525

## RESULT 6

T31098

probable dextrantransferase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides

C:Species: Leuconostoc mesenteroides

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000

C:Accession: T31098

R:Monchois, V.; Remaud-Simeon, M.; Monsan, P.; Willemot, R.M.

FEMS Microbiol. Lett. 159, 307-315, 1998

A:Title: Cloning and sequencing of a gene coding for an extracellular dextrantransferase

A:Reference number: 220981; MUID:98164374

A:Accession: T31098

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1508 <MON>

A:Cross-references: EMBL:AF030129; NID:g2766611; PID:g2766612; PIDN:AAB95453.1

A:Experimental source: strain NRRL B-1299

C:Genetics:

A:Gene: dsrB

C:Function:



A:Residues: 1-1599 <JAC>  
A:Cross-references: EMBL:Z11872; NID:g47530; PIDN:CAA7898.1; PID:g47531  
A:Experimental source: ATCC 25975  
R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.  
J. Gen. Microbiol. 137, 2577-2593, 1991  
A:Title: Molecular characterization of a cluster of at least two glucosyltransferase genes  
A:Reference number: A44811; MUID:92148377  
A:Accession: S28810  
A:Molecule type: DNA  
A:Residues: 1-51 <GIF>  
A:Cross-references: EMBL:Z11873  
C:Genetics:  
A:Gene: gtfK  
C:Superfamily: cpl repeat homology  
C:Keywords: glycosyltransferase; hexosyltransferase  
F:1456-1475/Domain: cpl repeat homology <CPR>

Query Match 54.0%; Score 61; DB 2; Length 1599;  
Best Local Similarity 57.1%; Pred. No. 0.32;  
Matches 12; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 AIDHLSILEANSGNDNDYVKQ 21  
I: ::||| ||| ||| :  
DB 491 ALANISILEANSHNDPPYVNE 511

RESULT 12  
B841176  
A:Title: Hypothetical protein Vng0156c [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: B841176  
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablon, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Liu  
A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483  
A:Accession: B841176  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-232 <STO>  
A:Cross-references: GB:AE004437; NID:g10579804; PIDN:AAG18774.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: VNG0156C

Query Match 43.4%; Score 49; DB 2; Length 232;  
Best Local Similarity 47.4%; Pred. No. 2.9;  
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 AIDHLSILEANSGNDNDYV 19  
I: | : ::||| | | :  
DB 62 AADPVRSVDANSGRDADHV 80

RESULT 13  
ISBYT2  
A:Title: Topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) - yeast (Saccharomyces cerevisiae)  
N:Alternate names: DNA gyrase; DNA topoisomerase II; protein N2244; protein YNL088W  
C:Species: Saccharomyces cerevisiae  
C:Date: 30-Sep-1992 #sequence\_revision 10-Nov-1995 #text\_change 16-Jun-2000  
C:Accession: S57534; A25630; S63027; S30866; S65093  
R:Soler-Mira, A.; Saiz, J.E.; Ballesta, J.P.G.; Remacha, M.  
submitted to the EMBL Data Library, June 1995  
A:Reference number: S57533  
A:Accession: S57534  
A:Molecule type: DNA  
A:Residues: 1-1428 <SOL>  
A:Cross-references: EMBL:X89016; NID:g887621; PIDN:CAA61422.1; PID:g887623  
R:Gisveer, G.; Lynn, R.; Goto, T.; Wang, J.C.

J. Biol. Chem. 261, 12448-12454, 1986  
A:Title: The complete nucleotide sequence of the structural gene TOP2 of yeast DNA to  
A:Reference number: A25630; MUID:86304413  
A:Accession: A25630  
A:Molecule type: DNA  
A:Residues: 1-74, 'N', '75-546, 'L', '548-836, 'R', '838-1428 <GIA>  
A:Cross-references: GB:M3814; NID:g172997; PIDN:AAB36610.1; PID:g172998  
R:Soler-Mira, A.; Saiz, J.E.; Ballesta, J.P.G.; Remacha, M.  
submitted to the Protein Sequence Database, April 1996  
A:Reference number: S63018  
A:Accession: S63027  
A:Molecule type: DNA  
A:Residues: 1-1428 <SOW>  
A:Cross-references: EMBL:Z71364; NID:gl301988; PIDN:CAA95964.1; PID:gl301989; GSPDB:G  
R:Jannatipour, M.; Liu, Y.X.; Nitiss, J.L.  
submitted to the EMBL Data Library, January 1993  
A:Description: The top2-5 mutant of yeast topoisomerase II encodes an enzyme resistant  
A:Reference number: S30866  
A:Accession: S30866  
A:Molecule type: DNA  
A:Residues: 812-836, 'R', '838-882, 'P', '884, 'II', '887-977 <JAN>  
A:Cross-references: EMBL:L08968; NID:g172999; PIDN:AAB59328.1; PID:g173000  
R:Soler-Mira, A.; Saiz, J.E.; Ballesta, J.P.G.; Remacha, M.  
Yeast 12, 485-491, 1996  
A:Title: The sequence of a 17 933 bp segment of Saccharomyces cerevisiae chromosome X  
A:Reference number: S65092; MUID:96310628  
A:Accession: S65093  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-1428 <SOF>  
A:Cross-references: EMBL:X89016; NID:g887621; PIDN:CAA61422.1; PID:g887623  
C:Comment: Type II DNA topoisomerase catalyzes the ATP-dependent transient breakage,  
C:Genetics:  
A:Gene: SGD:TOP2; TOR3; TRF3; TOP2-5; MIPS:YNL088W  
A:Cross-references: SGD:S0005032; MIPS:YNL088W  
A:Map position: 14L  
C:Superfamily: eukaryotic type II DNA topoisomerase; phage T4 DNA topoisomerase (ATP-  
C:Keywords: ATP; DNA binding; DNA replication; heterotetramer; isomerase; phosphot

Query Match 43.4%; Score 49; DB 1; Length 1428;  
Best Local Similarity 50.0%; Pred. No. 23;  
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 4 HLSILEANSGNDNDYV 19  
I: | : ||| ||| :  
DB 623 HLKIFHSLOGNDKQVI 638

RESULT 14  
T04254  
A:Title: Hypothetical protein F20B18.100 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 30-Apr-1999  
C:Accession: T04254  
R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hoheisel, J.; Mewes, H.W.; Mayer, K.  
submitted to the Protein Sequence Database, March 1999  
A:Reference number: Z15263  
A:Accession: T04254  
A:Molecule type: DNA  
A:Residues: 1-378 <BEV>  
A:Cross-references: EMBL:AL049483  
A:Experimental source: cultivar Columbia; BAC clone F20B18  
C:Genetics:  
A:Map position: 4.  
A:Introns: 326/3  
A:Note: F20B18.100

Query Match 42.5%; Score 48; DB 2; Length 378;  
Best Local Similarity 41.2%; Pred. No. 7.3;  
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

```
QY      2 IDHLSILEAWSGNDNDY 18
      :|: :| | | | | : : :
Db     295 LDYDGVLEAWSGKESPF 311
```

**RESULT 15**

B72129  
probable outer membrane leader peptide (omp) CPn0020 - Chlamydoiphila pneumoniae (strain  
C; Species: Chlamydoiphila pneumoniae, Chlamydia pneumoniae  
C; Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000  
C; Accession: B72129  
R; Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
Nature Genet. 21: 385-389, 1999  
A; Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A; Reference number: A72000; MUID:99206606  
A; Accession: B72129  
A; Status: preliminary  
A; Molecule type: DNA  
A; Residues: 1-695 <ARN>  
A; Cross-references: GB:AE001587; GB:AE001363; NID:g4376271; PIDN:AAD18173.1; PID:g4376271  
Experimental source: strain CWL029  
Genetics:  
A; Gene: CPn0020

Query Match 42.5%; Score 48; DB 2; Length 695;  
Best Local Similarity 57.1%; Pred. NO. 15;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy	3	DHLSILEAWSGNDN	16
		:	
Db	562	DHMNRWEWIGNDN	575

Search completed: March 27, 2002, 14:01:20  
Job time: 482 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:27:04 ; Search time 53.4 Seconds  
(without alignments)  
14.419 Million cell updates/sec

Title: US-09-290-049a-13  
Perfect score: 113  
Sequence: 1 AIDHLSILEANGNDYVKQ 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

1 number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	108	95.6	1365	1	GTFS_STRDO
2	75	66.4	1462	1	GTFS_STRMU
3	71	64.6	1476	1	GTFC_STRMU
4	71	62.8	1375	1	GTFC_STRMU
5	62	54.9	1592	1	GTFC_STRDO
6	62	54.9	1592	1	GTFC_STRDO
7	49	43.4	1428	1	TOP2_YEAST
8	48	42.5	3206	1	POLG_PSBMV
9	47	41.6	680	1	YH2X_SCHPO
10	45	39.8	2366	1	TOXB_CLODI
11	44	38.9	341	1	ACOB_BACSU
12	44	38.9	487	1	STR4_HUMAN
13	43	38.1	572	1	PGMU_DICDI
14	43	38.1	674	1	COAA_BACTJ
15	42.5	37.6	69	1	Y029_BUCAI
16	42	37.2	295	1	VENB_VIBU
17	41.5	36.7	128	1	Y532_BUCAI
18	41.5	36.7	151	1	Y347_METJA
19	41.5	36.7	680	1	NCPR_CANTR
20	41.5	36.7	844	1	PHSG_DROME
21	41	36.3	116	1	Y960_HABIN
22	41	36.3	666	1	POL_FMDV
23	41	36.3	801	1	PIPA_DICDI
24	41	36.3	1634	1	DPOL_METJA
25	40.5	35.8	310	1	PIRL_ARATH
26	40.5	35.8	833	1	CA41_YEAST
27	40.5	35.8	1461	1	TOP2_CANAL
28	40	35.4	53	1	Y0RU_TTV1
29	40	35.4	121	1	RL5_SOLME
30	40	35.4	179	1	FLJA_SALAE
31	40	35.4	179	1	FLJA_SALAE
32	40	35.4	179	1	FLJA_SALAE
33	40	35.4	302	1	CHI4_SOLTU

34	40	35.4	316	1	CHI2_SOLTU
35	40	35.4	318	1	CHI3_SOLTU
36	40	35.4	321	1	CHI1_THEC
37	40	35.4	324	1	CHI2_TOBAC
38	40	35.4	328	1	CHI1_SOLTU
39	40	35.4	329	1	CHI1_TOBAC
40	40	35.4	329	1	YQGI_CAPEL
41	40	35.4	366	1	ALF2_CAPEL
42	40	35.4	366	1	DSRB_ARCFU
43	40	35.4	430	1	SERC_SPIOL
44	40	35.4	462	1	PPB3_BACSU
45	40	35.4	576	1	CBPS_YEAST

ALIGNMENTS

RESULT 1					
GTFS_STRDO	ID	GTFS_STRDO	STANDARD;	PRT;	1365 AA.
AC	P29336;				
DT	01-DEC-1992 (Rel. 24, Created)				
DT	01-DEC-1992 (Rel. 24, Last sequence update)				
DT	01-APR-1993 (Rel. 25, Last annotation update)				
DE	GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)				
DE	(SUCROSE 6-GLUCOSYLTRANSFERASE).				
GN	GTFS.				
OS	Streptococcus downei (Streptococcus sobrinus).				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;				
OC	Streptococcus.				
OX	NCBI_TaxID=1317;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MFE28;				
RX	MEDLINE=90316665; PubMed=2142479;				
RA	Gilmore K.S., Russell R.R., Ferretti J.J.;				
RT	"Analysis of the Streptococcus downei gtfs gene, which specifies a				
RL	glucosyltransferase that synthesizes soluble glucans.";				
RL	Infect. Immun. 58:2452-2458(1990).				
CC	-1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT				
CC	TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE				
CC	OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE				
CC	AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.				
CC	-1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =				
CC	D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).				
CC	-1- ENZYME REGULATION: GLUCAN SYNTHESIS BY GTF-S IS INDEPENDENT OF				
CC	PRIMER GLUCAN UNLIKE GTF-I.				
CC	-1- DISEASE: DENTAL CARIES.				
CC	-1- MISCELLANEOUS: GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA				
CC	1,6-GLUCOSE).				
CC	-1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-				
CC	BINDING PROTEIN FROM S.MUTANS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; M30943; AAA26898.1; ..				
DR	PIR; A41483; A41483.				
DR	InterPro; IPR002479; CW_binding.				
DR	InterPro; IPR003318; Glyco_hydro_70.				
DR	Pfam; PF01473; CW_binding_1; 10.				
DR	Pfam; PF02324; Glyco_hydro_70; 1.				
KW	Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.				
FT	SIGNAL 1 36				
FT	OR 37 (POTENTIAL).				
FT	CHAIN 37 1365				
FT	GLUCOSYLTRANSFERASE-S.				
FT	CATALYTIC (APPROXIMATE).				
FT	DOMAIN 37 1050				
FT	GLUCAN-BINDING (APPROXIMATE).				
FT	DOMAIN 1083 1365				
FT	4.5 X TANDEM REPEATS.				



```

RESULT 3
GTFB_STRMU STANDARD: PRT: 1476 AA.
AC P08987; 069381; 069384; 069387; 069390; 069396;
DT 01-NOV-1988 (Rel. 09, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTFB.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";
J. Bacteriol. 169:4263-4270(1987).
[2]
SEQUENCE FROM N.A.
RC STRAIN=MT4239, MT4245, MT4251, MT4467, AND MT8148;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -!- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) =
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- DISEASE: DENTAL CARIES.
CC -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S. MUTANS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; M17361; AAA8588.1; -
CC EMBL; D88651; BAA26101.1; -
CC EMBL; D88654; BAA26105.1; -
CC EMBL; D88657; BAA26109.1; -
CC EMBL; D88660; BAA26113.1; -
CC EMBL; D89977; BAA26119.1; -
CC PIR; B33135; B33135.
CC InterPro; IPR002479; CW_binding.
CC InterPro; IPR003318; Glyco_hydro_70.
CC Pfam; PF01473; CW_binding_1; 13.
CC Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1 34 POTENTIAL.
FT CHAIN 35 1476 GLUCOSYLTRANSFERASE-I.
FT DOMAIN 35 1051 CATALYTIC (APPROXIMATE).
FT DOMAIN 1097 1476 GLUCAN-BINDING (APPROXIMATE).
FT REPEAT 1097 1130 A REPEAT.
FT DOMAIN 1161 1470 5 X TANDEM REPEATS.
FT REPEAT 1161 1210 1.
FT REPEAT 1225 1275 2.

```

```

FT REPEAT 1290 1340 3.
FT REPEAT 1355 1405 4.
FT REPEAT 1420 1470 5.
FT VARIANT 62 65 S -> T (IN STRAIN MT4239).
FT VARIANT 65 65 MT4251, MT4467 AND MT8148.
FT VARIANT 68 68 I -> T (IN STRAINS MT4239, MT4245,
FT VARIANT 78 78 A -> V (IN STRAIN MT4251).
FT VARIANT 86 86 Q -> P (IN STRAIN MT4239).
FT VARIANT 88 86 S -> I (IN STRAIN MT4251).
FT VARIANT 89 89 S -> F (IN STRAIN MT4239).
FT VARIANT 168 168 S -> N (IN STRAIN MT4251).
FT VARIANT 276 276 K -> N (IN STRAINS MT4239, MT4245 AND
FT VARIANT 399 399 D -> S (IN STRAINS MT4239,
FT VARIANT 474 474 N -> R (IN STRAIN MT4239).
FT VARIANT 512 512 I -> T (IN STRAIN MT4239).
FT VARIANT 519 519 K -> R (IN STRAIN MT8148).
FT VARIANT 701 701 F -> Y (IN STRAIN MT8148).
FT VARIANT 708 708 T -> I (IN STRAIN MT8148).
FT VARIANT 938 938 A -> V (IN STRAIN MT8148).
FT VARIANT 952 957 F -> L (IN STRAIN MT8148).
FT VARIANT 963 964 YGTPVA -> FGKPE (IN STRAINS MT4245,
FT VARIANT 964 964 MT4251 AND MT8148).
FT VARIANT 968 970 NT -> SV (IN STRAINS MT4245, MT4251 AND
FT VARIANT 1086 1086 MT8148).
FT VARIANT 1158 1158 A -> T (IN STRAIN MT4239).
FT VARIANT 1163 1163 S -> N (IN STRAIN MT4239).
FT VARIANT 1168 1168 H -> Y (IN STRAIN MT4251).
FT VARIANT 1182 1182 E -> K (IN STRAIN MT8148).
FT VARIANT 1234 1234 Y -> C (IN STRAIN MT8148).
FT VARIANT 1263 1263 A -> P (IN STRAIN MT4239).
FT VARIANT 1263 1263 H -> P (IN STRAIN MT8148).
FT VARIANT 1263 1263 H -> R (IN STRAINS MT4239, MT4245 AND
FT VARIANT 1264 1264 MT4251).
FT VARIANT 1272 1272 H -> Y (IN STRAINS MT4245 AND MT4251).
FT VARIANT 1329 1329 G -> S (IN STRAINS MT4245 AND MT4251).
FT VARIANT 1329 1329 Y -> H (IN STRAINS MT4239, MT4245, MT4251
FT VARIANT 1394 1394 AND MT8148).
FT VARIANT 1402 1402 H -> Y (IN STRAINS MT4245 AND MT4251).
FT VARIANT 1459 1459 G -> S (IN STRAINS MT4245 AND MT4251).
FT CONFLICT 570 570 Y -> H (IN STRAIN MT4467).
FT CONFLICT 800 817 R -> A (IN REF. 1).
FT CONFLICT 817 817 ADODVRAASTAPSTDGK -> LTKMFALRLARPHQOMA
FT CONFLICT 1310 1310 (IN REF. 1).
FT CONFLICT 1476 1476 H -> L (IN REF. 1).
FT SEQUENCE 1476 AA; 165685 MW; 3479B62B07694D98 CRC64;

Query Match 64.6%; Score 73; DB 1; Length 1476;
Best Local Similarity 73.7%; Pred. No. 0.0025;
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Oy 1 AIDHLSILEAWSGNDNDYV 19
Db 481 ANDHLSILEAWSDNDTPYL 499

RESULT 4
GTFB_STRMU STANDARD: PRT: 1375 AA.
AC DT P13470; P05427;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI PRECURSOR (EC 2.4.1.5) (GTF-SI)
DE (DEXTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTFB.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]

```







```

Best Local Similarity 47.6%; Pred. No. 48;
Matches 10; Conservative 4; Mismatches 5; Indels 2; Gaps 1;

QY 2 IDHLSI--LEAWSGNDNDYV 20
    ||| || : : | : |||
Db 394 IDHFSIVGIKNWAFDAEYK 414

RESULT 9
YH2X_SCHPO
ID YH2X_SCHPO STANDARD; PRT; 680 AA.
AC OT4343; 2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 76.4 KDA PROTEIN CIA4.09 IN CHROMOSOME II.
GN SPBC1A4.09.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
NCBI_TaxID=4896;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Wood V., Rajandream M.A., Barrell B.G., Volkkaert G.;
RA Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE UPF0024 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL031174; CAA20114.1; -
DR InterPro; IPR001656; UPF0024.
DR Pfam; PF01142; UPF0024; 1.
KW Hypothetical protein.
SQ SEQUENCE 680 AA; 76438 MW; B4A52DAECEA1B9A6 CRC64;

Query Match 41.6%; Score 47; DB 1; Length 680;
Best Local Similarity 62.5%; Pred. No. 12;
Matches 10; Conservative 1; Mismatches 3; Indels 2; Gaps 1;

QY 6 SILEAW--SGNDNDYV 19
    |||| | ||| ||:
434 SILETWSRSGNQTDYL 449

RESULT 10
TOXB_CLODI
ID TOXB_CLODI STANDARD; PRT; 2366 AA.
AC P18177;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE TOXIN B.
GN TOXB OR TCDB.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OC NCBI_TaxID=1496;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RX MEDLINE=90326540; PubMed=2374729;
RA Barroso L.A., Wang S.Z., Phelps C.J., Johnson J.L., Wilkins T.D.;
RA "Nucleotide sequence of Clostridium difficile toxin B gene.";
RT Nucleic Acids Res. 18:4004-4004(1990).

```

```

[2]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RA von Eichel-Streiber C.;
RA Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE OF 1271-2366 FROM N.A.
RC STRAIN=VPI 10463;
RX MEDLINE=92293124; PubMed=1603068;
RA Eichel-Streiber C., Laufenberg-Feldmann R., Sartingen S.,
RA Schulze J., Sauerborn M.;
RA "Comparative sequence analysis of the Clostridium difficile toxins A
RA and B.";
RT Mol. Gen. Genet. 233:260-268(1992).
CC -1- DISEASE: CLOSTRIDIUM DIFFICILE PRODUCES TWO TOXINS, AN
CC ENTEROTOXIN CALLED A AND CYTOTOXIN B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X53138; CAA37298.1; -
DR EMBL; X92982; CAA63562.1; -
DR EMBL; X60984; CAA43299.1; -
DR PIR; S10317; S10317.
DR InterPro; IPR002479; CW_binding.
DR Pfam; PF01473; CW_binding_1; 18.
KW Cytotoxin; Toxin.
SQ SEQUENCE 2366 AA; 269709 MW; E1024BD8B8A56ADF CRC64;

Query Match 39.8%; Score 45; DB 1; Length 2366;
Best Local Similarity 44.4%; Pred. No. 99;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 IDHLSLEAWSGNDNDYV 19
    | : ||| || : ||| :
Db 1733 IDLSIRYVWSNGNDFI 1750

RESULT 11
ACOB_BACSU
ID ACOB_BACSU STANDARD; PRT; 341 AA.
AC O34591;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ACETOIN:2,6-DICHLOROPHENOLINDOPHENOL OXIDOREDUCTASE BETA SUBUNIT
DE (EC 1.1.1.-) (ACETOIN:DCPIP OXIDOREDUCTASE-BETA) (AO:DCPIP OR)
DE (TPP-DEPENDENT ACETOIN DEHYDROGENASE EI BETA-SUBUNIT).
GN ACOB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Huang M.;
RA Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=AC327;
RX MEDLINE=97124190; PubMed=8969503;
RA Yamamoto H., Uchiyama S., Sekiguchi J.;
RA "Cloning and sequencing of a 40.6 kb segment in the 73 degrees-76
RA degrees region of the Bacillus subtilis chromosome containing genes
RA for trehalose metabolism and acetoin utilization.";
RT Microbiology 142:3057-3065(1996).

```





use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; U61984; AB03667.1; -  
HSSP; P00949; 3PMG.  
DictyDB; DD05092; pgmA.  
InterPro; IPR001485; PGM\_PMM.  
Pfam; PF00408; PGM\_PMM; 1.  
PRINTS; PR00509; PGM\_PMM.  
PROSITE; PS00710; PGM\_PMM; 1.  
Isomerase; Phosphorylation; Magnesium.  
FORMS THE PHOSPHOSERINE INTERMEDIATE  
FT ACT\_SITE 120 120  
SQ SEQUENCE 572 AA; 63229 MW; ED7E8F58F16F563A CRC64;

Query Match 38.1%; Score 43; DB 1; Length 572;  
Best Local Similarity 31.6%; Pred. No. 40;  
Matches 6; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

1 AIDHLSILEAWSGNDNDYV 19

DB 167 SYDNLGLLTKTYEWDGEFV 185

#### RESULT 14

ID COAA\_BACTJ STANDARD; PRT; 674 AA.  
AC 087905;  
DT 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DE PESTICIDIAL CRYSTAL PROTEIN CRY24AA (INSECTICIDAL DELTA-ENDOTOXIN  
DE CRYXIVA(A)) (CRYSTALINE ENTOMOCIDAL PROTOXIN) (CRYSTAL PROTEIN)  
DE (INSECTICIDAL PROTEIN JEG72) (FRAGMENT).  
GN CRY24AA OR CRYXIVA(A).  
OS Bacillus thuringiensis (subsp. jegathesan).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=56955;

[1]

SEQUENCE FROM N.A.  
RA Kawalek M.D., Gill S.S.;  
RT "Isolation and characterization of insecticidal genes from Bacillus  
RT thuringiensis subsp. jegathesan."  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: PROMOTES COLLOIDSMOTIC LYSIS BY BINDING TO THE MIDGUT  
CC EPITHELIAL CELLS OF INSECTS.  
CC -!- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING  
CC SPOULATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART  
CC OF THE SPORE COAT.

CC -!- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE  
CC N-TERMINUS.  
CC -!- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; U88188; AAC61891.1; -  
DR InterPro; IPR001178; Endotoxin.  
DR Pfam; PF00555; endotoxin; 1.  
KW Toxin; sporulation.

NON\_TER 674 674

SQ SEQUENCE 674 AA; 75959 MW; DA3904DAB891C978 CRC64;

Query Match 38.1%; Score 43; DB 1; Length 674;  
Best Local Similarity 46.7%; Pred. No. 48;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 DHLISILEAWSGNDND 17

DB 138 DYLGALAWNKNKN 152

#### RESULT 15

Y029\_BUCAI  
ID Y029\_BUCAI STANDARD; PRT; 69 AA.  
AC P57141;  
DT 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL PROTEIN BU029.  
GN BU029.

OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum  
OS symbiotic bacterium).  
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.  
OX NCBI\_TaxID=118099;

RN [1]

SEQUENCE FROM N.A.

RC STRAIN-TOKYO 1998;

RX MEDLINE=20445173; PubMed=10993077;

RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;

RT "Genome sequence of the endocellular bacterial symbiont of aphids

RT Buchnera sp. APS."

RL Nature 407:81-86(2000).

CC -!- SIMILARITY: BELONGS TO THE COF/YBHA/YIDA/YIGL (E.COLI) / YCSE/YXEH

CC (B.SUBTILIS) FAMILY.

CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AP001118; BAB12756.1; -

DR InterPro; IPR000150; Hypotheset\_cof.

DR PROSITE; PS01228; COF\_1; FALSE\_NEG.

DR PROSITE; PS01229; COF\_2; FALSE\_NEG.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 69 AA; 7735 MW; 0020CC0D593EFD3 CRC64;

Query Match 37.6%; Score 42.5; DB 1; Length 69;

Best Local Similarity 52.6%; Pred. No. 4.2;

Matches 10; Conservative 3; Mismatches 3; Indels 3; Gaps 1;

QY 1 AIDHLSILEAWSGNDNDYV 19

DB 38 ALPHLEIIES---NDNDGV 53

Search completed: March 27, 2002, 14:27:05  
Job time: 1647 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:22:47 ; Search time 137.48 Seconds  
(without alignments)  
10.540 Million cell updates/sec

Title: US-09-290-049A-12

Perfect score: 119

Sequence: 1 ANNHVSIVEAWSNDTPYLHD 21

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 332938 seqs, 6899538 residues

Total number of hits satisfying chosen parameters: 332938

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_New:\*

1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep:\*

2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep:\*

3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep:\*

4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep:\*

5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep:\*

6: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep1:\*

7: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep:\*

8: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119	100.0	21	6	US-09-562-328-23
2	119	100.0	21	6	US-09-562-328-25
3	119	100.0	21	6	US-09-290-049A-12
4	119	100.0	21	6	US-09-290-049A-14
5	108	90.8	21	6	US-09-562-328-20
6	108	90.8	21	6	US-09-290-049A-1
7	108	90.8	21	6	US-09-562-328-21
8	99	83.2	21	6	US-09-562-328-22
9	74	62.2	21	6	US-09-290-049A-11
10	74	62.2	21	6	US-09-604-957-4
11	74	62.2	545	6	US-09-604-957-5
12	65	54.6	523	6	US-09-562-328-24
13	62	52.1	21	6	US-09-290-049A-13
14	62	52.1	21	6	US-09-604-957-6
15	55	46.2	584	6	US-09-614-150-4881
16	50	42.0	1252	6	US-09-604-957-7
17	48	40.3	535	6	US-09-604-957-8
18	48	40.3	1278	6	US-09-888-615-73
19	46.5	39.1	1604	6	US-09-675-784A-8751
20	46	38.7	127	6	US-09-614-150-12219
21	46	38.7	880	6	US-09-603-703B-2226
22	45	37.8	199	6	US-09-708-427-53175
23	45	37.8	638	6	US-09-708-427-53175
24	45	37.8	765	6	US-09-708-427-53175
25	45	37.8	947	6	US-09-708-427-53175

26 45 37.8 972 6 US-09-924-154-16 Sequence 16, Appl

27 44 37.0 15 7 US-10-067-649-90 Sequence 90, Appl

28 43.5 36.6 491 1 PCT-US02-03987-10190 Sequence 10190, A

29 43.5 36.6 491 1 PCT-US02-03987-13756 Sequence 13756, A

30 43.5 36.6 491 6 US-09-815-242-10190 Sequence 10190, A

31 43.5 36.6 491 6 US-09-815-242-13756 Sequence 13756, A

32 43.5 36.6 491 7 US-10-072-851-10190 Sequence 10190, A

33 43.5 36.6 491 7 US-10-072-851-13756 Sequence 13756, A

34 43 36.1 67 7 US-10-050-704-324 Sequence 324, Appl

35 43 36.1 87 7 US-10-050-704-149 Sequence 149, Appl

36 43 36.1 89 7 US-10-015-127-12494 Sequence 12494, A

37 43 36.1 304 5 US-09-623-791-83 Sequence 83, Appl

38 43 36.1 434 6 US-09-708-427-12697 Sequence 12697, A

39 43 36.1 461 6 US-09-708-427-12696 Sequence 12696, A

40 43 36.1 473 6 US-09-708-427-12695 Sequence 12695, A

41 43 36.1 2603 6 US-09-897-516-5891 Sequence 5891, Ap

42 42.5 35.7 1059 6 US-09-614-150-1785 Sequence 1785, Ap

43 42 35.3 179 6 US-09-675-784A-8401 Sequence 8401, Ap

44 42 35.3 226 6 US-09-708-427-53175 Sequence 53175, A

45 42 35.3 254 8 US-60-356-051-2856 Sequence 2856, Ap

## ALIGNMENTS

RESULT 1

US-09-562-328-23

; Sequence 23, Application US/09562328

; GENERAL INFORMATION:

; APPLICANT: LEES, ANDREW

; APPLICANT: TAUBMAN, MARTIN A.

; APPLICANT: SMITH, DANIEL J.

; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES

; FILE REFERENCE: 04995.0046-01

; CURRENT APPLICATION NUMBER: US/09/562,328

; CURRENT FILING DATE: 2000-05-01

; PRIOR APPLICATION NUMBER: 09/288,965

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 44

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 23

; LENGTH: 21

; TYPE: PRT

; ORGANISM: Streptococcus sp.

US-09-562-328-23

Query Match 100.0%; Score 119; DB 6; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1.6e-12;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANNHVSIVEAWSNDTPYLHD 21

Db 1 ANNHVSIVEAWSNDTPYLHD 21

RESULT 2

US-09-562-328-25

; Sequence 25, Application US/09562328

; GENERAL INFORMATION:

; APPLICANT: LEES, ANDREW

; APPLICANT: TAUBMAN, MARTIN A.

; APPLICANT: SMITH, DANIEL J.

; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES

; FILE REFERENCE: 04995.0046-01

; CURRENT APPLICATION NUMBER: US/09/562,328

; CURRENT FILING DATE: 2000-05-01

; PRIOR APPLICATION NUMBER: 09/288,965

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 44

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 25

; LENGTH: 21



```

RESULT 7
US-09-290-049A-10
; Sequence 10, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290,049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. mutans
9-290-049A-10

```

```

Query Match 90.8%; Score 108; DB 6; Length 21;
Best Local Similarity 85.7%; Pred. No. 8.8e-11;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 ANNHVSIVEAWSNDNTPYLHD 21
      |||:|||||:|||||:|||||
Db 1 ANDHLSILEAWSNDNTPYLHD 21

```

```

RESULT 8
US-09-562-328-21
; Sequence 21, Application US/09562328
; GENERAL INFORMATION:
; APPLICANT: LEES, ANDREW
; APPLICANT: TAUBMAN, MARTIN A.
; APPLICANT: SMITH, DANIEL J.
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES
; FILE REFERENCE: 04995.0046-01
; CURRENT APPLICATION NUMBER: US/09/562,328
; CURRENT FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 09/288,965
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Streptococcus sp.
US-09-562-328-21

```

```

Query Match 83.2%; Score 99; DB 6; Length 21;
Best Local Similarity 81.0%; Pred. No. 2.4e-09;
Matches 17; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 ANNHVSIVEAWSNDNTPYLHD 21
      |||:|||||:|||||:|||||
Db 1 ANDHLSILEAWSNDNTPYLHD 21

```

```

RESULT 9
US-09-562-328-22
; Sequence 22, Application US/09562328
; GENERAL INFORMATION:
; APPLICANT: LEES, ANDREW
; APPLICANT: TAUBMAN, MARTIN A.
; APPLICANT: SMITH, DANIEL J.
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES
; FILE REFERENCE: 04995.0046-01

```

```

; CURRENT APPLICATION NUMBER: US/09/562,328
; CURRENT FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 09/288,965
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Streptococcus sp.
US-09-562-328-22

```

```

Query Match 62.2%; Score 74; DB 6; Length 21;
Best Local Similarity 66.7%; Pred. No. 2.3e-05;
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

```

```

Qy 1 ANNHVSIVEAWSNDNTPYLHD 21
      |||:|||||:|||||:|||||
Db 1 AINHLSILEAWSNDNTPQYNKD 21

```

```

RESULT 10
US-09-290-049A-11
; Sequence 11, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290,049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. mutans
US-09-290-049A-11

```

```

Query Match 62.2%; Score 74; DB 6; Length 21;
Best Local Similarity 66.7%; Pred. No. 2.3e-05;
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

```

```

Qy 1 ANNHVSIVEAWSNDNTPYLHD 21
      |||:|||||:|||||:|||||
Db 1 AINHLSILEAWSNDNTPQYNKD 21

```

```

RESULT 11
US-09-604-957-4
; Sequence 4, Application US/09604957
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 545

```

; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-604-957-4

Query Match 62.2%; Score 74; DB 6; Length 545;  
Best Local Similarity 66.7%; Pred. No. 0.00099;  
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ANNHVSIVEAWSNDNTPYLHD 21  
Db 75 AINHLSEAWSNDNPQYVKD 95  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 12  
US-09-604-957-5  
; Sequence 5, Application US/09604957  
; GENERAL INFORMATION:  
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
; APPLICANT: DIJKHUIZEN, LUBBERT  
; APPLICANT: RAHAOUI, HAKIM  
; APPLICANT: LEER, ROBERT-JAN  
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
; FILE REFERENCE: BO 43388  
; CURRENT APPLICATION NUMBER: US/09/604,957  
; CURRENT FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 00201871.1  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 523  
; TYPE: PRT  
; ORGANISM: Leuconostoc mesenteroides  
US-09-604-957-5

Query Match 54.6%; Score 65; DB 6; Length 523;  
Best Local Similarity 57.1%; Pred. No. 0.026;  
Matches 12; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 ANNHVSIVEAWSNDNTPYLHD 21  
Db 75 ANQHLSELDWSHNDPLXYVD 95  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 13  
US-562-328-24  
; Sequence 24, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT APPLICATION NUMBER: US/09/562,328  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 24  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-24

Query Match 52.1%; Score 62; DB 6; Length 21;  
Best Local Similarity 57.9%; Pred. No. 0.0019;  
Matches 11; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ANNHVSIVEAWSNDNTPYL 19

Db 1 AIDHLSILEAWSGNDNDYV 19  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 14  
US-09-290-049A-13  
; Sequence 13, Application US/09290049A  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: S. downei  
US-09-290-049A-13

Query Match 52.1%; Score 62; DB 6; Length 21;  
Best Local Similarity 57.9%; Pred. No. 0.0019;  
Matches 11; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ANNHVSIVEAWSNDNTPYL 19  
Db 1 AIDHLSILEAWSGNDNDYV 19  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 15  
US-09-604-957-6  
; Sequence 6, Application US/09604957  
; GENERAL INFORMATION:  
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
; APPLICANT: DIJKHUIZEN, LUBBERT  
; APPLICANT: RAHAOUI, HAKIM  
; APPLICANT: LEER, ROBERT-JAN  
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
; FILE REFERENCE: BO 43388  
; CURRENT APPLICATION NUMBER: US/09/604,957  
; CURRENT FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 00201871.1  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 584  
; TYPE: PRT  
; ORGANISM: Leuconostoc mesenteroides  
US-09-604-957-6

Query Match 46.2%; Score 55; DB 6; Length 584;  
Best Local Similarity 45.0%; Pred. No. 1.1;  
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Qy 1 ANNHVSIVEAWSNDNTPYLH 20  
Db 75 ANKHLSELDWNGKDPQYVN 94  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Search completed: March 27, 2002, 14:22:47  
Job time: 1694 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:01:19 ; Search time 102.51 Seconds  
(without alignments)  
15.605 Million cell updates/sec

Title: US-09-290-049a-12  
Perfect score: 119  
Sequence: 1 ANNHVSIVEAWSNDTPLYLHD 21  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Maximum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	119	100.0	1592	2 A38175	glucosyltransferas
2	108	90.8	1475	2 B33135	gtfB protein precu
3	99	83.2	1375	2 JT0345	dextranucrase (EC
4	74	62.2	1431	2 A45866	probable dextranu
5	68	57.1	1508	2 T31098	glucosyltransferas
6	67	56.3	1365	2 A41483	glucosyltransferas
7	61	51.3	1449	2 T30857	glucosyltransferas
8	61	51.3	1449	2 T30552	glucosyltransferas
9	57	47.9	1577	2 T30858	glucosyltransferas
10	56	47.1	1518	2 A44811	glucosyltransferas
11	55	46.2	1599	2 S22737	glucosyltransferas
12	53	44.5	175	2 C86205	hypothetical prote
13	49	41.2	336	1 QOBB40	BGLF2 protein - hu
14	49	41.2	524	2 D82220	conserved hypothet
15	48	40.3	331	2 B48445	glyceraldhyde-3-p
16	46	38.7	313	2 S59448	hypothetical prote
17	46	38.7	490	2 H70338	probable ppdk prot
18	45	37.8	337	2 C64233	glyceraldhyde-3-p
19	45	37.8	347	2 T48610	hypothetical prote
20	45	37.8	525	2 T40088	RhoGEF domain cont
21	45	37.8	947	2 E86362	hypothetical prote
22	45	37.8	1384	2 T02748	hypothetical prote
23	45	37.8	4848	2 T30289	pristinamycin I sy
24	44.5	37.4	835	2 B64689	site-specific DNA-
25	44	37.0	741	2 JC5142	X-Pro dipeptidyl-p
26	44	37.0	1122	2 S64443	probable membrane
27	44	37.0	1131	2 T16217	hypothetical prote
28	44	37.0	1313	2 T29037	hypothetical prote
29	43.5	36.6	491	1 D64947	glucose-6-phosphat

30	43.5	36.6	491	2 F85797	glucose-6-phosphat
31	43	36.1	78	2 S76593	hypothetical prote
32	43	36.1	194	2 H72037	conserved hypothet
33	43	36.1	194	2 C86586	CT647 hypothetical
34	43	36.1	275	2 S55978	hypothetical prote
35	43	36.1	284	2 JC7148	heat-shock sigma f
36	43	36.1	331	1 DEUTGC	glyceraldhyde-3-p
37	43	36.1	346	2 E70715	hypothetical prote
38	43	36.1	396	2 T35024	probable glutathio
39	43	36.1	441	2 T36737	hypothetical prote
40	43	36.1	453	2 T16795	hypothetical prote
41	43	36.1	473	2 T04218	hypothetical prote
42	43	36.1	504	2 C86428	hypothetical prote
43	43	36.1	528	2 E71525	probable ABC trans
44	43	36.1	603	2 S03020	dopamine beta-mono
45	43	36.1	899	2 B48586	suppressor of hair

ALIGNMENTS

RESULT 1  
A38175 glucosyltransferase precursor - Streptococcus sobrinus  
C:Species: Streptococcus sobrinus  
C:Date: 28-Aug-1992 #sequence\_revision 28-Aug-1992 #text\_change 15-Oct-1999  
C:Accession: A38175  
R:Abou, H.; Msumura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.  
J. Bacteriol. 173, 989-996, 1991  
A:Title: Peptide sequences for sucrose splitting and glucan binding within Streptococ  
A:Reference number: A38175; MUID:91123227  
A:Accession: A38175  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1592 <ABO>  
A:Cross-references: GB:D90213; NID:g217032; PIDN:BAA14241.1; PID:d1014946; PID:g21703  
C:Superfamily: cpl repeat homology  
F:1093-1112/Domain: cpl repeat homology <CP1>  
F:1222-1241/Domain: cpl repeat homology <CP2>  
F:1287-1306/Domain: cpl repeat homology <CP3>  
F:1330-1351/Domain: cpl repeat homology <CP4>  
F:1352-1371/Domain: cpl repeat homology <CP5>  
F:1402-1420/Domain: cpl repeat homology <CP6>  
F:1465-1484/Domain: cpl repeat homology <CP7>  
F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match 100.0%; Score 119; DB 2; Length 1592;  
Best Local Similarity 100.0%; Pred. No. 4.2e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANNHVSIVEAWSNDTPLYLHD 21  
|||||  
Db 477 ANNHVSIVEAWSNDTPLYLHD 497

RESULT 2  
B33135 gtfB protein precursor - Streptococcus mutans  
C:Species: Streptococcus mutans  
C:Date: 23-Oct-1990 #sequence\_revision 23-Oct-1990 #text\_change 15-Oct-1999  
C:Accession: B33135; A33128  
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
J. Bacteriol. 169, 4263-4270, 1987  
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.  
A:Reference number: A33135; MUID:87308013  
A:Accession: B33135  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1475 <SHI>  
A:Cross-references: GB:M17361; NID:g153639; PIDN:AAA88588.1; PID:g153640  
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
submitted to the Protein Sequence Database, September 1990

A:Reference number: A33128

A:Accession: A33128

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-171,173-641,'N',643-1475 <SH2>

A:Experimental source: strain GS-5

C:Superfamily: cpl repeat homology

F:1096-1115/Domain: cpl repeat homology <CP1>

F:1224-1243/Domain: cpl repeat homology <CP2>

F:1289-1308/Domain: cpl repeat homology <CP3>

F:1354-1373/Domain: cpl repeat homology <CP4>

F:1419-1438/Domain: cpl repeat homology <CP5>

Query Match

Best Local Similarity 90.8%; Score 108; DB 2; Length 1475;

Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 1 ANNHVIVEAWSNDNDTPYLHD 21

481 ANDHLSILEAWSNDNDTPYLHD 501

RESULT 3

JT0345

dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)

C:Species: Streptococcus mutans

C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 24-Sep-1999

C:Accession: JT0345; C33135

R:Ueda, S.; Shiroza, T.; Kuramitsu, H.K.

Gene 69, 101-109, 1988

A:Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.

A:Reference number: JT0345; MUID:89137980

A:Accession: JT0345

A:Molecule type: DNA

A:Residues: 1-1375 <UED>

A:Experimental source: GS-5

R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

J. Bacteriol. 169, 4263-4270, 1987

A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.

A:Reference number: A33135; MUID:87308013

A:Accession: C33135

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-349 <SHI>

Cross-references: GB:M17361

A:Gene: gtfC

C:Function:

A:Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans

C:Superfamily: cpl repeat homology

C:Keywords: duplication; glycosyltransferase; hexosyltransferase

F:1-34/Domain: signal sequence #status predicted <SIG>

F:35-1375/Product: glycosyltransferase #status predicted <MAT>

F:1126-1145/Domain: cpl repeat homology <CP1>

F:1253-1272/Domain: cpl repeat homology <CP2>

F:1318-1337/Domain: cpl repeat homology <CP3>

Query Match

Best Local Similarity 83.2%; Score 99; DB 2; Length 1375;

Matches 17; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ANNHVIVEAWSNDNDTPYLHD 21

507 ANDHLSILEAWSNDNDTPYLHD 527

RESULT 4

A45866

dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans

C:Species: Streptococcus mutans

C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999

C:Accession: A45866

R:Honda, O.; Kato, C.; Kuramitsu, H.K.

J. Gen. Microbiol. 136, 2099-2105, 1990

A:Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the gluco

A:Reference number: A45866; MUID:91100958

A:Accession: A45866

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1431 <HON>

A:Cross-references: GB:M29296

C:Superfamily: cpl repeat homology

C:Keywords: glycosyltransferase; hexosyltransferase

F:181-201/Domain: cpl repeat homology <CP1>

F:1127-1146/Domain: cpl repeat homology <CP2>

F:1192-1211/Domain: cpl repeat homology <CP3>

F:1257-1276/Domain: cpl repeat homology <CP4>

F:1277-1297/Domain: cpl repeat homology <CP5>

F:1321-1340/Domain: cpl repeat homology <CP6>

F:1341-1361/Domain: cpl repeat homology <CP6>

F:1385-1404/Domain: cpl repeat homology <CP7>

Query Match

Best Local Similarity 62.2%; Score 74; DB 2; Length 1431;

Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ANNHVIVEAWSNDNDTPYLHD 21

495 AINHLSILEAWSNDNDPQYNKD 515

RESULT 5

T31098

probable dextranucrase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides

C:Species: Leuconostoc mesenteroides

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000

C:Accession: T31098

R:Monchois, V.; Renaud-Simeon, M.; Monsan, P.; Willemot, R.M.

FEMS Microbiol. Lett. 159, 307-315, 1998

A:Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase

A:Reference number: Z20981; MUID:98164374

A:Accession: T31098

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1508 <MON>

A:Cross-references: EMBL:AF030129; NID:g2766611; PID:g2766612; PIDN:AAB95453.1

A:Experimental source: strain NRRL B-1299

C:Genetics:

A:Gene: dsrB

C:Function:

A:Description: produces dextran composed only of alpha(1-6) glucosidic bonds

C:Keywords: glycosyltransferase; hexosyltransferase

Query Match

Best Local Similarity 57.1%; Score 68; DB 2; Length 1508;

Matches 12; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 ANNHVIVEAWSNDNDTPYLHD 21

563 ANQHLSILEAWSNDNDPEVKD 583

RESULT 6

A41483

glycosyltransferase (EC 2.4.1.-) gtfS precursor - Streptococcus sobrinus

C:Species: Streptococcus sobrinus

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 15-Oct-1999

C:Accession: A41483

R:Gillmore, K.S.; Russell, R.R.B.; Ferretti, J.J.

Infect. Immun. 58, 2452-2458, 1990

A:Title: Analysis of the Streptococcus downei gtfS gene, which specifies a glucosyltr





A: Molecule type: DNA  
A: Residues: 1-331 <HAN>  
A: Cross-references: EMBL:X65220; NID:g9552; PIDN:CAA46323.1; PID:g9553  
C: Superfamily: glyceraldehyde-3-phosphate dehydrogenase  
C: Keywords: oxidoreductase

Query Match 40.3%; Score 48; DB 2; Length 331;  
Best Local Similarity 42.1%; Pred. No. 7.6;  
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
Qy 2 NNHVSIVEAKSDNDTPYLH 20  
| : : | : : |  
Db 301 NDHFVKLVSWYDNETGYSH 319

Search completed: March 27, 2002, 14:01:20  
Job time: 482 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:27:03 ; Search time 53.4 Seconds  
(without alignments)  
14.419 Million cell updates/sec

Title: US-09-290-049A-12  
Perfect score: 119  
Sequence: 1 ANHVSIVEAWSNDPTPLHD 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 segs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119	100.0	1592	1	GTF2_STRDO
2	119	100.0	1597	1	GTF1_STRDO
3	108	90.8	1476	1	GTFB_STRMU
4	99	83.2	1375	1	GTFB_STRMU
5	74	62.2	1462	1	GTFD_STRMU
6	67	56.3	1365	1	GTFB_STRDO
7	49	41.2	336	1	UL16_EBV
8	48	40.3	330	1	G3PC_LEIME
9	46	38.7	313	1	YM58_YEAST
10	45	37.8	337	1	G3PC_MYCGE
11	44	37.0	1122	1	YG3C_YEAST
12	44	37.0	1131	1	YPO4_CAEEL
13	44	37.0	3358	1	PGCV_MOUSE
14	43.5	36.6	491	1	G6PD_ECOLI
15	43	36.1	275	1	YG33_YEAST
16	43	36.1	330	1	G3PC_TRYBB
17	43	36.1	346	1	Y943_MYCTU
18	43	36.1	441	1	HGD_STRCO
19	43	36.1	603	1	DOPO_HUMAN
20	43	36.1	899	1	SUHW_DROVI
21	42.5	35.7	470	1	AMP2_STRCO
22	42.5	35.7	491	1	G6PD_ERWCH
23	42	35.3	279	1	CG16_YEAST
24	42	35.3	334	1	G3P_BACST
25	42	35.3	335	1	G3P_BACCO
26	42	35.3	432	1	HGD_PSEAE
27	42	35.3	775	1	AD28_HUMAN
28	42	35.3	776	1	AD28_MACFA
29	42	35.3	1627	1	ADP1_MYCPN
30	42	35.3	3898	1	POLG_HCVB
31	41.5	34.9	416	1	HGD_LEGPN
32	41.5	34.9	468	1	KG3H_DICDI
33	41.5	34.9	578	1	VAC8_YEAST

34 41 34.5 130 1 UCR6\_SCHMA  
35 41 34.5 236 1 UCR4\_TOBAC  
36 41 34.5 241 1 HFB1\_HABIN  
37 41 34.5 241 1 HFB2\_HABIN  
38 41 34.5 301 1 Y186\_MYCPN  
39 41 34.5 383 1 DUHM\_BSPFI  
40 41 34.5 621 1 DCTB\_RHIME  
41 41 34.5 1139 1 INAI\_CAEEL  
42 41 34.5 1634 1 DPOL\_METJA  
43 41 34.5 2485 1 PTND\_HUMAN  
44 41 34.5 3341 1 POLG\_MCFB  
45 41 34.5 3898 1 POLG\_BVDVS

#### ALIGNMENTS

#### RESULT 1

GTF2\_STRDO  
ID GTF2\_STRDO STANDARD; PRT: 1592 AA.  
AC P27470;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)  
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
OS Streptococcus downei (Streptococcus).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1317;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=6715;  
RX MEDLINE=91123227; PubMed=1704006;  
RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,  
RA Kagawa H.;  
RT "Peptide sequences for sucrose splitting and glucan binding within  
RT streptococcus sobrinus glucosyltransferase (water-insoluble glucan  
RT synthetase).";  
RL J. Bacteriol. 173:989-996(1991).  
CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -!- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =  
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
CC -!- SUBCELLULAR LOCATION: SECRETED.  
CC -!- DISEASE: DENTAL CARIES.  
CC -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
CC FORMS OF GLUCANS.  
CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
CC BINDING PROTEIN FROM S. MUTANS.  
-----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL: D90213; BAA14241.1;  
CC PIR: A38175; A38175.  
CC HSP: P00695; 2HEE.  
CC InterPro: IPR002479; CW\_binding.  
CC InterPro: IPR003318; Glyco\_hydro\_70.  
CC Pfam: PF01473; CW\_binding.1; 16.  
CC Pfam: PF02324; Glyco\_hydro\_70; 1.  
CC Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.  
FT SIGNAL 1 38  
FT CHAIN 39 1592 GLUCOSYLTRANSFERASE-I.

FT DOMAIN 39 1044 CATALYTIC (APPROXIMATE).  
 FT DOMAIN 1093 1592 GLUCAN-BINDING (APPROXIMATE).  
 FT DOMAIN 1093 1592 6.5 X TANDEM REPEATS.  
 FT REPEAT 1093 1142 1.  
 FT REPEAT 1158 1207 2.  
 FT REPEAT 1222 1272 3.  
 FT REPEAT 1287 1337 4.  
 FT REPEAT 1402 1451 5.  
 FT REPEAT 1514 1563 6.  
 FT REPEAT 1577 1592 7 (INCOMPLETE).  
 SQ SEQUENCE 1592 AA; 176167 MW; BC0A66D079351ECF CRC64;

Query Match 100.0%; Score 119; DB 1; Length 1592;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANNHVSIVEAWSNDTPYLHD 21  
 |||||  
 477 ANNHVSIVEAWSNDTPYLHD 497

## RESULT 2

GTFL\_STRDO STANDARD; PRT; 1597 AA.  
 AC P11001;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)  
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
 GN GTFI.  
 OS Streptococcus downei (Streptococcus sobrinus).  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1317;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MFE28;  
 RX MEDLINE=87308014; PubMed=3040686;  
 RA Ferretti J.J., Gilpin M.L., Russell R.R.B.;  
 RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus sobrinus MFE28";  
 RL J. Bacteriol. 169:4271-4278(1987).  
 CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -!- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) = D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -!- SUBCELLULAR LOCATION: SECRETED.  
 CC -!- DISEASE: DENTAL CARIES  
 CC -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF GLUCANS.  
 CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-BINDING PROTEIN FROM S.MUTANS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M17391; AAC63063.1; .  
 DR InterPro; IPR002479; CW binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 19.  
 DR Pfam; PF03324; Glyco\_hydro\_70; 1.  
 KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.

FT SIGNAL 1 38 POTENTIAL.  
 FT CHAIN 39 1597 GLUCOSYLTRANSFERASE-I.  
 FT DOMAIN 39 1050 CATALYTIC (APPROXIMATE).  
 FT DOMAIN 1099 1597 GLUCAN-BINDING (APPROXIMATE).  
 FT DOMAIN 1099 1597 1.25 A, 2 B, AND 5 AC REPEATS.  
 FT REPEAT 1099 1132 A REPEAT.  
 FT REPEAT 1163 1213 AC REPEAT.  
 FT REPEAT 1227 1277 AC REPEAT.  
 FT REPEAT 1292 1342 AC REPEAT.  
 FT REPEAT 1352 1399 B REPEAT.  
 FT REPEAT 1406 1455 AC REPEAT.  
 FT REPEAT 1465 1512 B REPEAT.  
 FT REPEAT 1519 1568 AC REPEAT.  
 FT REPEAT 1582 1597 A REPEAT (INCOMPLETE).  
 SQ SEQUENCE 1597 AA; 177080 MW; B9E86A200868798E CRC64;

Query Match 100.0%; Score 119; DB 1; Length 1597;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANNHVSIVEAWSNDTPYLHD 21  
 |||||  
 483 ANNHVSIVEAWSNDTPYLHD 503

## RESULT 3

GTFB\_STRMU STANDARD; PRT; 1476 AA.  
 ID AC P08987; 069381; 069384; 069387; 069390; 069396;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)  
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
 GN GTFB.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-GS-5;  
 RX MEDLINE=87308013; PubMed=3040685;  
 RA Shiroza T., Ueda S., Kuramitsu H.K.;  
 RT "Sequence analysis of the gtfB gene from Streptococcus mutans";  
 RL J. Bacteriol. 169:4263-4270(1987).  
 CC [2]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN-MT4239, MT4245, MT4251, MT4467, AND MT8148;  
 CC MEDLINE=98231643; PubMed=9570124;  
 CC Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S., Kimura S., Hamada S.;  
 RT "Molecular analyses of glucosyltransferase genes among strains of Streptococcus mutans";  
 CC FEMS Microbiol. Lett. 161:331-336(1998).  
 CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -!- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) = D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -!- SUBCELLULAR LOCATION: SECRETED.  
 CC -!- DISEASE: DENTAL CARIES.  
 CC -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF GLUCANS.  
 CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-BINDING PROTEIN FROM S.MUTANS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -



the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```
CC EMBL; M17361; AAA88588.1; -
CC EMBL; D88651; BAA26101.1; -
CC EMBL; D88654; BAA26105.1; -
CC EMBL; D88657; BAA26109.1; -
CC EMBL; D88660; BAA26113.1; -
CC EMBL; D89977; BAA26119.1; -
CC PIR; B33135;
CC InterPro; IPR002479; CW_binding.
CC InterPro; IPR003318; Glyco_hydro_70.
CC Pfam; PF01473; CW_binding_1; 13.
CC Pfam; PF02324; Glyco_hydro_70; 1.
CC Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.
CC SIGNAL 1 34
CC CHAIN 35 1476
CC DOMAIN 35 1051
CC REPEAT 1097 1476
CC REPEAT 1097 1130
CC DOMAIN 1161 1470
CC REPEAT 1161 1210
CC REPEAT 1225 1275
CC REPEAT 1290 1340
CC REPEAT 1355 1405
CC REPEAT 1420 1470
CC VARIANT 62 62
CC VARIANT 65 65
CC VARIANT 68 68
CC VARIANT 78 78
CC VARIANT 86 86
CC VARIANT 89 89
CC VARIANT 168 168
CC VARIANT 276 276
CC VARIANT 399 399
CC VARIANT 474 474
CC VARIANT 512 512
CC VARIANT 519 519
CC VARIANT 701 701
CC VARIANT 708 708
CC VARIANT 938 938
CC VARIANT 952 957
CC VARIANT 963 964
CC VARIANT 968 970
CC VARIANT 1086 1086
CC VARIANT 1158 1158
CC VARIANT 1163 1163
CC VARIANT 1168 1168
CC VARIANT 1182 1182
CC VARIANT 1234 1234
CC VARIANT 1263 1263
CC VARIANT 1263 1263
CC VARIANT 1264 1264
CC VARIANT 1272 1272
CC VARIANT 1329 1329
CC VARIANT 1394 1394
CC VARIANT 1402 1402
CC VARIANT 1459 1459
CC CONFLICT 570 570
CC CONFLICT 800 817
CC CONFLICT 1310 1310
CC SEQUENCE 1476 AA; 165685 MW; 3479B62B07694D98 CRC64;
```

Query Match 90.88; Score 108; DB 1; Length 1476;  
Best Local Similarity 85.7%; Pred. No. 5.5e-09;  
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANNEHSIVEAWSNDNDTPYLHD 21  
DB 481 ANDHLSILEAWSNDNDTPYLHD 501  
||:|||||

## RESULT 4

```
GTFC_STRMU ID GTFC_STRMU STANDARD; PRT; 1375 AA.
AC P13470; P05427;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI PRECURSOR (EC 2.4.1.5) (GTF-SI)
DE (DEXTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTFC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=89137980; PubMed=2976010;
RA Ueda S., Shiroza T., Kuramitsu H.K.;
RT "Sequence analysis of the gtfc gene from Streptococcus mutans GS-5.";
RL Gene 69:101-109(1988).
RN [2]
RP SEQUENCE OF 1-349 FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfc gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) =
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DISEASE: DENTAL CARIES.
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S.MUTANS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

EMBL; M22054; AAA88592.1; -

EMBL; M17361; AAA88589.1; -

PIR; JTO345; JTO345.

PIR; C33135; C33135.

InterPro; IPR002479; CW\_binding.

InterPro; IPR003318; Glyco\_hydro\_70.

Pfam; PF01473; CW\_binding\_1; 7.

Pfam; PF02324; Glyco\_hydro\_70; 1.

Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.

SIGNAL 1 34

CHAIN 35 1375

GLUCOSYLTRANSFERASE-SI.







```
DR EMBL; U02213; AAD12507.1; -.
DR EMBL; U02178; AAD12463.1; -.
DR HSP; P17721; 1HDG.
DR TIGR; MG301; -.
DR InterPro; IPR000173; GAP_DH.
DR Pfam; PF00044; gpdh; 1.
DR PRINTS; PR00078; G3PDHGRGNASE.
DR PROSITE; PS00071; GAPDH; 1.
KW Glycolysis; Oxidoreductase; NAD; Complete proteome.
FT BINDING 157 157
FT GLYCERALDEHYDE 3-PHOSPHATE
FT ACT_SITE 184 184
FT ACTIVATES THIOL GROUP DURING CATALYSIS
FT (BY SIMILARITY).
FT (BY SIMILARITY).
SQ SEQUENCE 337 AA; 37097 MW; FALEA1966687006B CRC64;

Query Match 37.8; Score 45; DB 1; Length 337;
Best Local Similarity 53.8; Pred. No. 9.1;
Matches 7; Conservative 3; Mismatches 0; Indels 3; Gaps 0;

8 YEASDNDTPYLH 20
I I I I I I I I
313 VYWDNESSIVH 325

RESULT 11
YG3C_YEAST STANDARD; PRT; 1122 AA.
AC P53280;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 128.8 KDA PROTEIN IN PAS2-PRS5 INTERGENIC REGION.
GN YGRL34W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA van Dyck L., Skala J., de Wergifosse P., Purnelle B., Talla E.,
RA Nawrocki A., del Bino S., Goffeau A.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; 272919; CAA97147.1; -.
DR SGD; S0003366; YGRL34W.
KW Hypothetical protein.
SQ SEQUENCE 1122 AA; 128779 MW; BF3542816CC78490 CRC64;

Query Match 37.0%; Score 44; DB 1; Length 1122;
Best Local Similarity 46.7%; Pred. No. 55;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 ANNHVSIVEAWSND 15
DQ 110 SSSYASIVLSWKND 124

RESULT 12
YPO4_CAEEL STANDARD; PRT; 1131 AA.
ID YPO4_CAEEL
AC YP5015;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 127.2 KDA PROTEIN F31E3.4 IN CHROMOSOME III.
GN F31E3.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Pavello A.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO YEAST PAN2 AND TO S-POMBE SPAC22G7.04.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U40935; AAA81687.1; -.
DR Wormpep; F31E3.4; CE01269.
DR InterPro; IPR000520; Exonuclease.
DR InterPro; IPR001394; UCH-2.
DR Pfam; PF000929; Exonuclease; 1.
DR Pfam; PF004400; WD40; 3.
DR SMART; SM00479; EXOIII; 1.
DR Hypothetical protein.
KW Hypothetical protein.
SQ SEQUENCE 1131 AA; 127174 MW; DESA4AB0C309E7DE CRC64;

Query Match 37.0%; Score 44; DB 1; Length 1131;
Best Local Similarity 36.8%; Pred. No. 56;
Matches 7; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 3 NHVSIVEAWSNDTPYLHD 21
DQ 315 NHCGLVNVFADRDQPVNE 333

RESULT 13
PGCV_MOUSE STANDARD; PRT; 3358 AA.
ID PGCV_MOUSE
AC Q62059; Q62058;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VERSICAN CORE PROTEIN PRECURSOR (LARGE FIBROBLAST PROTEOGLYCAN)
DE (CHONDROITIN SULFATE PROTEOGLYCAN CORE PROTEIN 2) (PG-M).
DE CSPG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (VARIANTS V0; V1 AND V2).
RC STRAIN=C57BL/6, AND SWISS WEBSTER; TISSUE=Brain;
RX MEDLINE=95122551; PubMed=7822336;
RA Ito K., Shinomura T., Zako M., Ujita M., Kimata K.;
RT "Multiple forms of mouse PG-M, a large chondroitin sulfate
RT proteoglycan generated by alternative splicing.";
RL J. Biol. Chem. 270:958-965(1995).
RN [2]
RP SEQUENCE OF 1-348 AND 3053-3358 FROM N.A. (VARIANT V3).
RC STRAIN=C57BL/6;
RX MEDLINE=95181355; PubMed=7876137;
RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;
RT "Expression of PG-M(V3), an alternatively spliced form of PG-M
RT without a chondroitin sulfate attachment in region in mouse and human
RT tissues.";
```

J. Biol. Chem. 270:3914-3918(1995).

-1- FUNCTION: MAY PLAY A ROLE IN INTERCELLULAR SIGNALING AND IN CONNECTING CELLS WITH THE EXTRACELLULAR MATRIX. MAY TAKE PART IN THE REGULATION OF CELL MOTILITY, GROWTH AND DIFFERENTIATION. BINDS HYALURONIC ACID.

-1- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX.

-1- ALTERNATIVE PRODUCTS: AT LEAST 4 ISOFORMS; V0 (SHOWN HERE), V1, V2 AND V3; ARE PRODUCED BY ALTERNATIVE SPLICING.

-1- TISSUE SPECIFICITY: V2 IS FOUND ONLY IN BRAIN.

-1- DEVELOPMENTAL STAGE: DISAPPEARS AFTER THE C-YPE DEVELOPMENT.

-1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.

-1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.

-1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.

-1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.

-1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.

-----

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; D16263; BAA03796.1; -

EMBL; D28599; -; NOT ANNOTATED\_CDS.

EMBL; D32040; BAA06802.1; -

HSP; P00740; IIXA.

MGI; MGI:102889; Cspg2.

InterPro; IPR000152; Asx\_hydroxyl.

InterPro; IPR000561; EGF-like.

InterPro; IPR000742; EGF-2.

InterPro; IPR001881; EGF\_Ca.

InterPro; IPR001438; EGF\_II.

InterPro; IPR003599; Ig.

InterPro; IPR003006; Ig\_MHC.

InterPro; IPR000538; Link.

InterPro; IPR000436; Sush1\_SCR\_CCP.

InterPro; IPR001304; lectin\_c.

Pfam; PF00008; EGF; 2.

Pfam; PF00047; Ig; 1.

Pfam; PF00059; lectin\_c; 1.

Pfam; PF00084; sush1; 1.

Pfam; PF00193; Xlink; 2.

PRINTS; PR00010; EGFBL00D.

ProDom; PD000918; Link; 2.

SMART; SM00032; CCP; 1.

SMART; SM00034; CLECT; 1.

SMART; SM00179; EGF\_CA; 1.

SMART; SM00001; EGF-like; 1.

SMART; SM00409; IG; 1.

SMART; SM00445; LINK; 2.

PROSITE; PS00010; ASX\_HYDROXYL; 1.

PROSITE; PS00022; EGF-1; 2.

PROSITE; PS01186; EGF-2; 1.

PROSITE; PS01187; EGF\_Ca; 1.

PROSITE; PS01241; LINK; 2.

PROSITE; PS00615; C-TYPE LECTIN-1; 1.

PROSITE; PS00041; C-TYPE LECTIN-2; 1.

Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sush1; Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain; Hyaluronic acid; Alternative splicing.

SIGNAL 1 20 POTENTIAL.

CHAIN 21 3358 VERSICAN CORE PROTEIN.

DOMAIN 37 137 IG-LIKE V-TYPE DOMAIN.

DOMAIN 167 244 LINK 1.

DOMAIN 265 334 LINK 2.

DOMAIN 348 1308 GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN).

DOMAIN 1309 3052 GAG-BETA.

DOMAIN 3052 3088 EGF-LIKE 1.

DOMAIN 3090 3126 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).

FT	DOMAIN	3129	3255	C-TYPE LECTIN.
FT	DOMAIN	3258	3316	SUSHI.
FT	DISULFID	44		BY SIMILARITY.
FT	DISULFID	172	243	BY SIMILARITY.
FT	DISULFID	196	217	BY SIMILARITY.
FT	DISULFID	270	333	BY SIMILARITY.
FT	DISULFID	294	315	BY SIMILARITY.
FT	DISULFID	3056	3067	BY SIMILARITY.
FT	DISULFID	3061	3076	BY SIMILARITY.
FT	DISULFID	3078	3087	BY SIMILARITY.
FT	DISULFID	3094	3105	BY SIMILARITY.
FT	DISULFID	3099	3114	BY SIMILARITY.
FT	DISULFID	3116	3125	BY SIMILARITY.
FT	DISULFID	3132	3143	BY SIMILARITY.
FT	DISULFID	3160	3252	BY SIMILARITY.
FT	DISULFID	3228	3244	BY SIMILARITY.
FT	DISULFID	3259	3302	BY SIMILARITY.
FT	DISULFID	3288	3315	BY SIMILARITY.
FT	CARBOHYD	57	57	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	330	330	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	351	351	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	441	441	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	807	807	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	914	914	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	951	951	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1305	1305	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1372	1372	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1679	1679	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2054	2054	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2244	2244	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2362	2362	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2627	2627	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3030	3030	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3332	3332	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3342	3342	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARSPIC	349	1308	MISSING (IN ISOFORM V1).
FT	VARSPIC	349	3052	MISSING (IN ISOFORM V2).
FT	CONFLICT	349	3052	MISSING (IN ISOFORM V3).
FT	CONFLICT	348	348	P -> R (IN REF. 2).
SEQ	SEQUENCE	3358 AA; 366938 MW; 071B800265C0762D CRC64;		

Query Match 37.0%; Score 44; DB 1; Length 3358;  
 Best Local Similarity 50.0%; Pred. No. 2e+02;  
 Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 NHVSIVEAWSNDTPY 18  
 |||:| | | | |  
 Db 510 NHISLKELPEKNKTPY 525

RESULT 14  
 G6PD\_ECOLI  
 ID G6PD\_ECOLI STANDARD; PRT; 491 AA.  
 AC P22992; P78069; Q60134; Q60139;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE (EC 1.1.1.49) (G6PD).  
 GN ZWF OR B1852.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia  
 OC NCBI\_TaxID=562;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=K12;  
 RX MEDLINE=91123224; PubMed=1704005;  
 RA Rowley D.L., Wolf R.E. Jr.;  
 RT "Molecular characterization of the Escherichia coli K-12 zwf gene  
 RL encoding glucose 6-phosphate dehydrogenase.";  
 RL J. Bacteriol. 173:968-977(1991).  
 RN [2]

SEQUENCE FROM N.A.  
RP STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of *Escherichia coli* K-12.";  
RL Science 277:1453-1474(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX STRAIN=K12;  
RX MEDLINE=97251358; PubMed=9097040;  
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,  
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,  
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,  
RA Saito N., Sempel G., Seki Y., Sivasubramanian S., Tagami H.,  
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;  
RT "A 460-kb DNA sequence of the *Escherichia coli* K-12 genome  
corresponding to the 40.1-50.0 min region on the linkage map.";  
RN DNA Res. 3:379-392(1996).  
RX [4]  
RP SEQUENCE OF 72-368 FROM N.A.  
RX STRAIN=VARIOUS ECOR STRAINS;  
RX MEDLINE=95064015; PubMed=7973728;  
RA Guttman D.S., Dykhuizen D.E.;  
RT "Clonal divergence in *Escherichia coli* as a result of recombination,  
RT not mutation.";  
RL Science 266:1380-1383(1994).  
RN [5]  
RP SEQUENCE OF 321-491 FROM N.A.  
RX STRAIN=K12;  
RX MEDLINE=93345818; PubMed=8344525;  
RA Carter A.T., Pearson B.M., Dickinson J.R., Lancashire W.E.;  
RT "Sequence of the *Escherichia coli* K-12 edd and eda genes of the  
RT Entner-Doudoroff pathway.";  
RL Gene 130:155-156(1993).  
RN [6]  
RP -1- CATALYTIC ACTIVITY: D-GLUCOSE 6-PHOSPHATE + NADPH.  
RX DELTA-LACTONE 6-PHOSPHATE + NADPH.  
CC -1- PATHWAY: FIRST STEP IN PENTOSE PHOSPHATE PATHWAY.  
CC -1- SIMILARITY: BELONGS TO THE GLUCOSE-6-PHOSPHATE DEHYDROGENASE  
CC FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M55005; AAA24775.1; -;  
DR EMBL; AE000279; AAC74922.1; -;  
DR EMBL; D90828; BAA15660.1; -;  
DR EMBL; U13783; AAA57018.1; -;  
DR EMBL; U13784; AAA57019.1; -;  
DR EMBL; U13785; AAA57020.1; -;  
DR EMBL; U13786; AAA57021.1; -;  
DR EMBL; U13787; AAA57022.1; -;  
DR EMBL; U13788; AAA57023.1; -;  
DR EMBL; U13789; AAA57024.1; -;  
DR EMBL; U13790; AAA57025.1; -;  
DR EMBL; U13791; AAA57026.1; -;  
DR EMBL; U13792; AAA57027.1; -;  
DR EMBL; U13793; AAA57028.1; -;  
DR EMBL; U13794; AAA57029.1; -;  
DR EMBL; X63694; CAA45220.1; -;  
DR PIR; A38174; A38174.  
DR HSSP; P11411; 2DPG.  
DR SWISS-2DPAGE; P22992; COLI.  
DR ECO2DBASE; F048.8; 6TH EDITION.  
DR EcoGene; Egl11221; zwf.

DR InterPro; IPR001282; G6PD.  
DR Pfam; PF00479; G6PD; 1.  
DR PRINTS; PR00079; G6PDHGRNASE.  
DR PRODOM; PD001129; G6PD; 1.  
DR PROSITE; PS00069; G6P\_DEHYDROGENASE; 1.  
KW Oxidoreductase; NADP; Glucose metabolism; Complete proteome.  
FT ACT\_SITE 181 BY SIMILARITY.  
FT VARIATION 100 S -> N (IN STRAINS ECOR4 AND ECOR10).  
FT CONFLICT 268 LKSLRRIDRSNVREKTVRGQYTAGFA -> PEVSSPHRPLQ  
FT CONFLICT 293 RTRKTRTRAIYCV (IN REF. 1).  
SQ SEQUENCE 491 AA; 55704 MW; 263F07D298EAFCD3 CRC64;  
  
Query Match 36.6%; Score 43.5; DB 1; Length 491;  
Best Local Similarity 69.2%; Pred. No. 25;  
Matches 9; Conservative 1; Mismatches 2; Indels 1; Gaps 1;  
  
Qy 6 SIVEAWS-DNDTP 17  
Db 451 SITEAWAMDNDAP 463  
  
RESULT 15  
YG33\_YEAST STANDARD; PRT; 275 AA.  
AC P53271;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE HYPOTHETICAL 31.8 KDA PROTEIN IN NUP57-MEPI INTERGENIC REGION.  
GN YGR120C OR G6324.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID:4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=S288C / FY1679;  
RX MEDLINE=97197982; PubMed=9046098;  
RA van Dyck L., Tettelin H., Purnelle B., Goffeau A.;  
RT "An 18.3 kb DNA fragment from yeast chromosome VII carries four  
RT of Ty and three tRNA genes.";  
RL Yeast 13:171-176(1997).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X83099; CAA58155.1; -;  
DR EMBL; Z72905; CAA97130.1; -;  
DR SGD; S0003352; SEC35.  
KW Hypothetical protein.  
SQ SEQUENCE 275 AA; 31799 MW; AA102D086FF3FAD7 CRC64;  
  
Query Match 36.1%; Score 43; DB 1; Length 275;  
Best Local Similarity 42.9%; Pred. No. 15;  
Matches 6; Conservative 7; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 7 IVEAWSNDTPYLH 20  
Db 217 ILESCADNSPYIH 230  
  
Search completed: March 27, 2002, 14:27:04  
Job time: 1646 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:26:04 ; Search time 188.53 Seconds  
(without alignments) updates/sec  
16.293 Million cell

Title: US-09-290-049a-12  
Perfect score: 119  
Sequence: 1 ANNHVSIVEAWSNDPTPLD 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
al number of hits satisfying chosen parameters: 473505

Maximum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

SPTREMBL\_17:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119	100.0	1590	2 Q55263	Q55263 streptococc
2	119	100.0	1590	2 Q59983	Q59983 streptococc
3	108	90.8	1390	2 Q69385	Q69385 streptococc
4	108	90.8	1455	2 Q69391	Q69391 streptococc
5	99	83.2	1455	2 Q69388	Q69388 streptococc
6	99	83.2	1455	2 Q69388	Q69388 streptococc
7	99	83.2	1455	2 Q69397	Q69397 streptococc
8	76	63.9	1016	2 Q9LCJ7	Q9LCJ7 leuconostoc
9	68	57.1	1477	2 Q9L466	Q9L466 leuconostoc
10	68	57.1	1508	2 Q52224	Q52224 leuconostoc
11	68	57.1	1508	2 Q9EZH5	Q9EZH5 leuconostoc
12	68	57.1	1575	2 Q9LCH3	Q9LCH3 streptococc
13	68	57.1	1577	2 Q54178	Q54178 streptococc
14	67	56.3	1338	2 Q9WXJ4	Q9WXJ4 streptococc
15	65	54.6	1527	2 Q9ZARA	Q9ZARA leuconostoc
16	61	51.3	1449	2 Q68542	Q68542 streptococc
17	61	51.3	1449	2 Q55264	Q55264 streptococc
18	60	50.4	1512	2 Q9WXJ5	Q9WXJ5 streptococc
19	57	47.9	1577	2 Q55265	Q55265 streptococc

20	56	47.1	1518	2 Q00600	Q00600 streptococc
21	55	46.2	1599	2 Q00599	Q00599 streptococc
22	55	46.2	2057	2 Q9RE05	Q9RE05 leuconostoc
23	53	44.5	175	10 Q9LWJ8	Q9LWJ8 arabisdopsis
24	51	42.9	1195	2 Q06459	Q06459 thermus aqu
25	50	42.0	1252	5 Q9VQR4	Q9VQR4 drosophila
26	49	41.2	524	2 Q9KSI8	Q9KSI8 vibrio chol
27	47	39.5	193	2 Q9S445	Q9S445 pseudomonas
28	46.5	39.1	176	2 Q9KK16	Q9KK16 yersinia en
29	46	38.7	271	11 Q9QZE9	Q9QZE9 rattus norv
30	46	38.7	490	2 Q06579	Q06579 mycobacteri
31	46	38.7	601	2 Q05566	Q05566 mycobacteri
32	46	38.7	880	5 P91643	P91643 drosophila
33	46	38.7	880	5 Q9VK54	Q9VK54 drosophila
34	45	37.8	172	10 Q9FJ22	Q9FJ22 arabisdopsis
35	45	37.8	347	10 Q9LY92	Q9LY92 arabisdopsis
36	45	37.8	525	3 Q59679	Q59679 schizosacch
37	45	37.8	947	10 Q23136	Q23136 arabisdopsis
38	45	37.8	1366	4 Q9V269	Q9V269 homo sapien
39	45	37.8	1384	4 Q75257	Q75257 homo sapien
40	45	37.8	1683	5 Q9NKM1	Q9NKM1 leishmania
41	45	37.8	3895	12 Q65464	Q65464 border dise
42	45	37.8	3896	12 P87514	P87514 pestivirus
43	45	37.8	4848	2 Q07944	Q07944 streptomyc
44	44.5	37.4	835	2 Q25908	Q25908 helicobacte
45	44	37.0	136	12 Q9Q714	Q9Q714 classical s

#### ALIGNMENTS

#### RESULT 1

Q55263 ID Q55263 PRELIMINARY; PRT; 1590 AA.  
AC Q55263;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-JUN-2001 (TREMBLrel. 05, Last sequence update)  
DE GLUCOSYLTRANSFERASE GTF-I.  
OS Streptococcus sobrinus.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
CC Streptococcus.  
OX NCBI\_TaxID=1310;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 33478;  
RA Sato S.;  
RL Ann. Kagoshima Univ. Dental School 16:23-29(1996).  
DR EMBL; D63570; BAA09792.1; -  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 15.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1590 AA; 176057 MW; 9DF7A3F2C6E4FD43 CRC64;

Query Match 100.0%; Score 119; DB 2; Length 1590;  
Best Local Similarity 100.0%; Pred. No. 1.1e-09;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

#### Qy 1 ANNHVSIVEAWSNDPTPLD 21

|||||  
Db 477 ANNHVSIVEAWSNDPTPLD 497

#### RESULT 2

Q59983 ID Q59983 PRELIMINARY; PRT; 1590 AA.  
AC Q59983;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-JUN-2001 (TREMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

```
DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE
DE 6-GLUCOSYLTRANSFERASE).
GN GTFI.
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OM2176;
RX MEDLINE=94146405; PubMed=8312602;
RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;
RT "DNA sequence of the glucosyltransferase gene of serotype d
RL Streptococcus sobrinus.";
RL DNA Seq. 4:19-27(1993).
CC -1- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) = D-
CC FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).
EMBL; D13858; BAA02976.1; -
InterPro; IPR002479; CW_binding.
InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 16.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Signal; Transferase; Glycosyltransferase.
FT SIGNAL 1 38 POTENTIAL.
FT CHAIN 39 1590 GLUCOSYLTRANSFERASE-I.
SQ SEQUENCE 1590 AA; 175955 MW; C3C83A57CF3C2B0E CRC64;

Query Match 100.0%; Score 119; DB 2; Length 1590;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANNHVSIWEASNDNDTPYLHD 21
|||||
Db 477 ANNHVSIWEASNDNDTPYLHD 497

RESULT 3
O69385 PRELIMINARY; PRT; 1390 AA.
ID O69385;
AC O69385;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTFC.
Streptococcus mutans.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT4245;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RL Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D88655; BAA26106.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 7.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1390 AA; 155375 MW; 8847E495EF05E9F CRC64;

Query Match 90.8%; Score 108; DB 2; Length 1390;
Best Local Similarity 85.7%; Pred. No. 4.7e-08;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANNHVSIWEASNDNDTPYLHD 21
|||||
Db 477 ANNHVSIWEASNDNDTPYLHD 497

RESULT 3
O69385 PRELIMINARY; PRT; 1390 AA.
ID O69385;
AC O69385;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTFC.
Streptococcus mutans.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT4245;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RL Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D88655; BAA26106.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 7.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1390 AA; 155375 MW; 8847E495EF05E9F CRC64;

Query Match 90.8%; Score 108; DB 2; Length 1390;
Best Local Similarity 85.7%; Pred. No. 4.7e-08;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANNHVSIWEASNDNDTPYLHD 21
|||||
Db 477 ANNHVSIWEASNDNDTPYLHD 497

RESULT 3
O69385 PRELIMINARY; PRT; 1390 AA.
ID O69385;
AC O69385;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTFC.
Streptococcus mutans.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT4245;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RL Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D88655; BAA26106.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 7.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1390 AA; 155375 MW; 8847E495EF05E9F CRC64;

Query Match 90.8%; Score 108; DB 2; Length 1590;
Best Local Similarity 85.7%; Pred. No. 4.9e-08;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ANNHVSIWEASNDNDTPYLHD 21
|||||
Db 507 ANNHVSIWEASNDNDTPYLHD 527

RESULT 5
O69382 PRELIMINARY; PRT; 1455 AA.
ID O69382;
AC O69382;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTFC.
Streptococcus mutans.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT8148;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RL Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D88652; BAA26102.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 9.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1455 AA; 162969 MW; 27D4D3A1EECA2939 CRC64;
```

```
Query Match      83.2%; Score 99; DB 2; Length 1455;
Best Local Similarity 81.0%; Pred. No. 1.2e-06;
Matches 17; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ANNHVSIVEAWSNDNDTPYLHD 21
   ||:|||||
Db 507 ANDHLSILEAWSNDNDTPYLHD 527

RESULT 6
O69388 PRELIMINARY; PRT; 1455 AA.
ID AC O69388;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE DE GLUCOSYLTRANSFERASE-SI.
GN GTF.
Streptococcus mutans.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT4239;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR -EMBL: D88658; BAA26110.1; -.
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1; 10.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1455 AA; 163045 MW; 6D90A4978D35DD82 CRC64;

Query Match      83.2%; Score 99; DB 2; Length 1455;
Best Local Similarity 81.0%; Pred. No. 1.2e-06;
Matches 17; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ANNHVSIVEAWSNDNDTPYLHD 21
   ||:|||||
Db 507 ANDHLSILEAWSNDNDTPYLHD 527

RESULT 7
O69397 PRELIMINARY; PRT; 1455 AA.
ID AC O69397;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE DE GLUCOSYLTRANSFERASE-SI.
GN GTF.
Streptococcus mutans.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT4467;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RL FEMS Microbiol. Lett. 161:331-336(1998).

Query Match      83.2%; Score 99; DB 2; Length 1455;
Best Local Similarity 81.0%; Pred. No. 1.2e-06;
Matches 17; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ANNHVSIVEAWSNDNDTPYLHD 21
   ||:|||||
Db 507 ANDHLSILEAWSNDNDTPYLHD 527

RESULT 8
O9LCJ7 PRELIMINARY; PRT; 1016 AA.
ID AC O9LCJ7;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE DE DEXTRANSUCRASE.
GN DSRT.
Leuconostoc mesenteroides.
Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL B-512F;
RX MEDLINE=20169623; PubMed=10705445;
RA Funane K., Mizuno K., Takahara H., Kobayashi M.;
RT "Gene encoding a dextranucrase-like protein in Leuconostoc
RL mesenteroides NRRL B-512F.";
RL Biosci. Biotechnol. Biochem. 64:29-38(2000).
DR EMBL: AB020020; BAA90527.1; -.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1016 AA; 110343 MW; 8896EFDE13CCCB47 CRC64;

Query Match      63.9%; Score 76; DB 2; Length 1016;
Best Local Similarity 61.9%; Pred. No. 0.0029;
Matches 13; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ANNHVSIVEAWSNDNDTPYLHD 21
   :|||:|
Db 548 SNQHVSILEDSDNDAEYVKD 568

RESULT 9
O9L466 PRELIMINARY; PRT; 1477 AA.
ID AC O9L466;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE DE DEXTRANSUCRASE (EC 2.4.1.5).
GN DSRC.
Leuconostoc mesenteroides.
Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL B-1355;
RA Arguello-Morales M.A., Remaud-Simeon M., Pizzut S., Sarcabal P.,
RA Willemot R.M., Monsan P.;
RT "Sequence analysis of the gene encoding alternansucrase, a sucrose
```

RT glucosyltransferase from *Leuconostoc mesenteroides* NRRL B-1355.<sup>+</sup>;  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AJ250172; CAB76565.1; -;  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 14.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase; Glycosyltransferase.  
 SQ SEQUENCE 1477 AA; 164886 MW; E6F5710DEDFCB831 CRC64;

Query Match 57.1%; Score 68; DB 2; Length 1477;  
 Best Local Similarity 57.1%; Pred. No. 0.076;  
 Matches 12; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 ANNHVSIIVEAWSNDNDTPYLHD 21  
 |||:|||||:|||||:|  
 532 ANQHLSILEDWSHNDPEYVKD 552

RESULT 10  
 O52224

ID O52224 PRELIMINARY; PRT; 1508 AA.

AC O52224;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE 6-  
 DE GLUCOSYLTRANSFERASE).  
 GN DSRB.

OS *Leuconostoc mesenteroides*.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC *Leuconostoc*.  
 OX NCBI\_TaxID=1245;  
 RN [1]

SEQUENCE FROM N.A.  
 RC STRAIN=NRRL B-1299;

RA Monchois V., Renaud-Simeon M., Monsan P., Willemot R.M.;  
 RL FEMS Microbiol. Lett. 0:0-0(1998).

CC -1- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) = D-  
 CC FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).  
 DR EMBL: AF050129; AAB95453.1; -;

DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 14.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.

DR Transferase; Glycosyltransferase.  
 SQ SEQUENCE 1508 AA; 168511 MW; E70CECB57A70D1F0 CRC64;

Query Match 57.1%; Score 68; DB 2; Length 1508;  
 Best Local Similarity 57.1%; Pred. No. 0.078;  
 Matches 12; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 ANNHVSIIVEAWSNDNDTPYLHD 21  
 |||:|||||:|||||:|  
 Db 563 ANQHLSILEDWSHNDPEYVKD 583

RESULT 11  
 Q9EZH5

ID Q9EZH5 PRELIMINARY; PRT; 1508 AA.

AC Q9EZH5;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE DEXTRANSUCRASE DSRB/42.  
 GN DSRB742.

OS *Leuconostoc mesenteroides*.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC *Leuconostoc*.  
 OX NCBI\_TaxID=1245;  
 RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=B-742CB;

RA Kim H.-S., Kim D., Ryu H.-J., Robyt J.F.;

RT "Leuconostoc mesenteroides B-742CB, a dextranucrase gene."  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF294469; AAG38021.1; -;

DR InterPro: IPR002479; CW\_binding.

DR InterPro: IPR003318; Glyco\_hydro\_70.

DR Pfam: PF01473; CW\_binding\_1; 14.

DR Pfam: PF02324; Glyco\_hydro\_70; 1.

SQ SEQUENCE 1508 AA; 168542 MW; E2FCFA0F87AE4F3A CRC64;

Query Match 57.1%; Score 68; DB 2; Length 1508;  
 Best Local Similarity 57.1%; Pred. No. 0.078;  
 Matches 12; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 ANNHVSIIVEAWSNDNDTPYLHD 21  
 |||:|||||:|||||:|  
 Db 563 ANQHLSILEDWSHNDPEYVKD 583

RESULT 12  
 Q9LCH3

ID Q9LCH3 PRELIMINARY; PRT; 1575 AA.

AC Q9LCH3;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE GLUCOSYLTRANSFERASE.

GN GTPF.

OS *Streptococcus oralis*.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC *Streptococcus*.

OX NCBI\_TaxID=1303;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC10557;

RX MEDLINE=20231779; PubMed=10768934;

RA Fujiwara T., Hoshino T., Sobue S., Hamada S.;

RT "Purification, characterization, and molecular analysis of the gene

RL encoding glucosyltransferase from *Streptococcus oralis*."

RL Infect. Immun. 68:2475-2483(2000).

DR EMBL: AB025228; BAA95201.1; -;

DR InterPro: IPR002479; CW\_binding.

DR InterPro: IPR003318; Glyco\_hydro\_70.

DR Pfam: PF01473; CW\_binding\_1; 17.

DR Pfam: PF02324; Glyco\_hydro\_70; 1.

KW Transferase.

SQ SEQUENCE 1575 AA; 176792 MW; 772A26E4D7C2E543 CRC64;

Query Match 57.1%; Score 68; DB 2; Length 1575;  
 Best Local Similarity 61.9%; Pred. No. 0.082;  
 Matches 13; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 ANNHVSIIVEAWSNDNDTPYLHD 21  
 |||:|||||:|||||:|  
 Db 546 AIKHLSEAWSNDPDYVKD 566

RESULT 13  
 Q54178

ID Q54178 PRELIMINARY; PRT; 1577 AA.

AC Q54178; Q54247;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE GLUCOSYLTRANSFERASE.

GN STFG.

OS *Streptococcus gordonii* Challis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC *Streptococcus*.

OX NCBI\_TaxID=29390;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CHALLIS;  
RX MEDLINE=96157084; PubMed=8586195;  
RA Vickerman M.M., Sulavik M.C., Clewell D.B.;  
RT "Molecular analysis of Streptococcus gordonii glucosyltransferase  
phase variants.";  
RL Dev. Biol. Stand. 85:309-314(1995).  
RN [2]  
RP SEQUENCE OF 1-96 FROM N.A.  
RC STRAIN=CHALLIS;  
RX MEDLINE=92276337; PubMed=1534326;  
RA Sulavik M.C., Tardif G., Clewell D.B.;  
RT "Identification of a gene, rgt, which regulates expression of  
glucosyltransferase and influences the Spp phenotype of Streptococcus  
gordonii Challis.";  
RL J. Bacteriol. 174:3577-3586(1992).  
DR EMBL; U12643; AAC43483.1; .  
DR EMBL; M89776; AAA26969.1; .  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 18.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1577 AA; 177805 MW; 5AE0328DC5E08D18 CRC64;  
  
Query Match 57.1%; Score 68; DB 2; Length 1577;  
Best Local Similarity 61.9%; Pred. No. 0.082;  
Matches 13; Conservative 2; Mismatches 6; Indels 0; Gaps 0;  
  
QY 1 ANNHVSIVEAWSNDNDTPYLHD 21  
| :||:||||| | | | | |  
Db 548 ALKHLISILEAWSNDNDPDYKND 568  
  
RESULT 14  
Q9WXJ4  
ID Q9WXJ4 PRELIMINARY; PRT; 1338 AA.  
AC Q9WXJ4;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE GTF-S.  
GN GTF-S.  
OS Streptococcus criceti.  
OG Plasmid pAM1.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
NCBI\_TaxID=1333;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HS-6;  
RA Inoue M., Fukui K., Miyagi A.;  
RT "S.cricetus glucosyltransferase(gtfS and gtfT) genes.";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBDJ databases.  
DR EMBL; AB026123; BAA77236.1; .  
DR HSSP; P06278; IVJS.  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 10.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Plasmid.  
SQ SEQUENCE 1338 AA; 148558 MW; 0A90C8E10E15D99B CRC64;

Query Match 56.3%; Score 67; DB 2; Length 1338;  
Best Local Similarity 57.1%; Pred. No. 0.098;  
Matches 12; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
  
QY 1 ANNHVSIVEAWSNDNDTPYLHD 21  
| :||:||||| | | | | |

Db 437 AIDHLSILEAWSGNDNDYVKD 457  
  
RESULT 15  
Q9ZAR4  
ID Q9ZAR4 PRELIMINARY; PRT; 1527 AA.  
AC Q9ZAR4;  
DT 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE DEXTRANSUCRASE.  
GN DEX.  
OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OC Leuconostoc.  
OX NCBI\_TaxID=1245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NRRL B-512-F;  
RA Bhatnagar R., Singh D.K.S.;  
RT "Cloning and Molecular Characterization of Dextranucrase Gene from  
Leuconostoc mesenteroides NRRL B-512F.";  
RL Submitted (DEC-1996) to the EMBL/GenBank/DBDJ databases.  
DR EMBL; U81374; AAD10952.1; .  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 16.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
SQ SEQUENCE 1527 AA; 169709 MW; 1DFAFA237C743398 CRC64;  
  
Query Match 54.6%; Score 65; DB 2; Length 1527;  
Best Local Similarity 57.1%; Pred. No. 0.23;  
Matches 12; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
  
QY 1 ANNHVSIVEAWSNDNDTPYLHD 21  
| :||:||||| | | | | |  
Db 581 ANQHLSILEDWSHNDPLXVTD 601

Search completed: March 27, 2002, 14:26:05  
Job time: 1677 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:57:56 ; Search time 198.55 Seconds  
(without alignments)  
7,834 Million cell updates/sec

Title: US-09-290-049a-13

Perfect score: 113

Sequence: 1 AIDHLSILEAWSGNDNDYVKQ 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_1101.\*  
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	73	64.6	1577	17 AAR91047	Alpha-D-glucosyltr
2	65	57.5	2057	21 AAB10667	L. mesenteroides a
3	62	54.9	1592	14 AAR32925	Glucosyltransferase
4	48.5	42.9	218	21 AAB43277	Human ORF3041
5	48.5	42.9	289	22 AAM25871	Human protein sequ
6	48.5	42.9	427	20 AAY25761	Human secreted pro
7	48.5	42.9	427	21 AAB32412	Human secreted pro
8	48.5	42.9	436	21 AAB32411	Human secreted pro
9	48.5	42.9	576	21 AAB32384	Human secreted pro
10	48.5	42.9	576	22 AAB94297	Human protein sequ
11	48	42.5	661	20 AAY34618	Chlamydia pneumoni

12	47	41.6	722	22	AAB82125	Alpha-1,3-multi-br
13	47	41.6	730	21	AAG28898	Arabidopsis thalia
14	47	41.6	748	21	AAG28897	Arabidopsis thalia
15	47	41.6	769	21	AAG28896	Arabidopsis thalia
16	46	40.7	1969	19	AAG72419	Rice bacterial lea
17	45	39.8	2366	17	AAR95011	C. difficile toxin
18	45	39.8	2366	19	AAW68388	Clostridium diffi
19	44	38.9	639	18	AAW26530	Trypanosoma cruzi
20	44	38.9	639	20	AAY23292	Trypanosoma cruzi
21	43	38.1	456	20	AAY02274	A F-box protein se
22	43	38.1	614	21	AAG39868	Arabidopsis thalia
23	43	38.1	659	21	AAG39867	Arabidopsis thalia
24	43	38.1	702	20	AAY36829	Chlamydia trachoma
25	42.5	37.6	499	22	AAB62218	Glycine max homogl
26	42	37.2	235	21	AAG33167	Zea mays protein f
27	42	37.2	240	22	AAG32594	C glutamicum prote
28	42	37.2	277	21	AAG33166	Zea mays protein f
29	42	37.2	291	21	AAG33165	Zea mays protein f
30	42	37.2	305	21	AAG09845	Arabidopsis thalia
31	42	37.2	339	21	AAG09844	Arabidopsis thalia
32	42	37.2	388	22	AAG91488	C glutamicum prote
33	42	37.2	487	20	AAY21674	Human Ste20 homolo
34	42	37.2	770	20	AAY43478	Amino acid sequenc
35	42	37.2	794	22	AAG90825	C glutamicum prote
36	42	37.2	1205	21	AAY49434	Blast disease-resi
37	42	37.2	1251	21	AAY93135	Rice blast fungus-
38	41.5	36.7	679	21	AAY90596	Candida tropicalis
39	41.5	36.7	679	21	AAY90597	Candida tropicalis
40	41.5	36.7	916	22	AAG92599	C glutamicum prote
41	41	36.3	132	19	AAW82564	Rice glycogenin cl
42	41	36.3	132	22	AAE05245	Rice glycogenin pr
43	41	36.3	224	21	AAW90789	M. jannaschii MJ08
44	41	36.3	492	22	AAB94170	Human protein sequ
45	41	36.3	610	22	AAB93588	Human protein sequ

#### ALIGNMENTS

#### RESULT 1

AAR91047 AAR91047 standard; Protein; 1577 AA.

XX AC AAR91047;

XX DT 22-MAY-1996 (first entry)

XX DE Alpha-D-glucosyltransferase.

XX KW Alpha-D-glucosyltransferase; primer-independent; soluble glucan;  
KW sucrose; transgenic plant; cloning; Escherichia coli;  
KW phage lambda-Cl3; vector; plasmid pSG501; plasmid pSG502;  
KW gene transfer; crop improvement; storage carbohydrate; pasture;  
KW feedstuff; senescence; dextran; binder; food; pharmaceutical.

XX OS Streptococcus salivarius strain ATCC 25975.

XX PN WO9606173-A1.

XX PD 29-FEB-1996.

XX PF 24-AUG-1995; 95WO-AU00527.

XX PR 24-AUG-1994; 94AU-0007643.

XX PA (GIFF/) GIFFARD P M.

XX PA (JACQ/) JACQUES N A.

XX PA (SIMP/) SIMPSON C L.

XX PI Giffard PM, Jacques NA, Simpson CL;

XX WP; 1996-151376/15.

XX DR N-PSDB; AAT13139.

XX Plants contg. new bacterial DNA encoding glucosyl transferase  
 PT activity - retain higher levels of stored carbohydrate(s) in a form  
 PT readily digestible by ruminants  
 XX  
 PS Claim 4; Page 16-20; 31pp; English.  
 XX  
 CC The sequence represents an alpha-D-glucosyltransferase from  
 CC Streptococcus salivarius. The enzyme is primer-independent, and  
 CC produces soluble glucan from sucrose. A gene encoding the enzyme  
 CC may be cloned and expressed in Escherichia coli using a subclone  
 CC of phage lambda-C13, e.g. plasmid pGS501 or plasmid pGS502. The  
 CC DNA may also be expressed in a transgenic plant, to improve the  
 CC level of stored carbohydrate in a pasture plant which normally  
 CC contains low levels, or to prevent degradation of stored carbohydrate  
 CC during plant senescence. Dextran may be isolated from the plant, for  
 CC use as a food binder or pharmaceutical additive. Primer independence  
 CC ensures that the enzyme will be functional in plants. The glucan is  
 CC poorly degraded in plants but easily degraded by bacteria in the rumen  
 CC of grazing livestock.  
 XX  
 SQ Sequence 1577 AA;

Query Match 64.6%; Score 73; DB 17; Length 1577;  
 Best Local Similarity 75.0%; Pred. No. 0.008;  
 Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 AIDHLSILEAWSGNDNDYVK 20  
 || ||||| ||: ||  
 Db 588 alahlsileawsyndhynk 607

## RESULT 2

AAB10667  
 ID AAB10667 standard; Protein; 2057 AA.  
 XX  
 AC AAB10667;  
 XX  
 DT 19-JAN-2001 (first entry)  
 XX  
 DE L. mesenteroides alternan sucrose protein.  
 XX  
 KW Alternan sucrose; glucosyltransferase; fructose; cosmetic; foodstuff;  
 KW syrup.  
 XX  
 XX Leuconostoc mesenteroides.  
 PA DE19905069-A1.  
 PD 10-AUG-2000.  
 XX  
 PF 08-FEB-1999; 99DE-1005069.  
 XX  
 PR 08-FEB-1999; 99DE-1005069.  
 XX  
 PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.  
 PA (PLAC) MAX PLANCK GES. FÖRDERUNG WISSENSCHAFTEN.  
 XX  
 PI Kossmann J, Welsh T, Quanz M, Knuth K;  
 XX  
 XX WPI; 2000-550294/51.  
 DR N-PSDB; AAA97904.  
 XX  
 XX New nucleic acid encoding recombinant Leuconostoc mesenteroides  
 PT alternan sucrose protein and methods of alternan and fructose  
 PT production  
 XX  
 PS Claim 1a; Page 30-36; 64pp; German.

XX This invention describes a novel nucleic acid molecule (I) encoding an  
 CC alternan sucrose (E.C. 2.4.1.140 - an enzyme, that belongs to the  
 CC glucosyltransferase group) The recombinant, purified alternan sucrose

CC gene is useful for the fermentative production of alternan (a  
 CC carbohydrate) and/or fructose by secreting the enzyme into a  
 CC saccharose-containing culture medium. Alternatively, the enzyme is  
 CC contacted with a saccharose-containing solution. The alternan and/or  
 CC fructose is then isolated from the medium. Cosmetic products or  
 CC foodstuffs containing alternan can be produced. Recombinant production of  
 CC alternan sucrose is advantageous as it provides a cost effective means of  
 CC producing fructose for high fructose containing syrups, production of  
 CC which previously has been achieved by costly production from maize  
 CC starch. This sequence represents the Leuconostoc mesenteroides alternan  
 CC sucrose protein which is described in the method of the invention.  
 XX  
 SQ Sequence 2057 AA;

Query Match 57.5%; Score 65; DB 21; Length 2057;  
 Best Local Similarity 66.7%; Pred. No. 0.21;  
 Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 4 HLSILEAWSGNDNDYVK 21  
 ||||| ||: ||  
 Db 668 hlsiledwngkdpqyvnq 685

## RESULT 3

AAR32925  
 ID AAR32925 standard; Protein; 1592 AA.  
 XX  
 AC AAR32925;  
 XX  
 DT 28-JUN-1993 (first entry)  
 XX  
 DE Glucosyltransferase I.  
 XX  
 KW GT-1; Streptococcus; dental; caries.  
 XX  
 OS Streptococcus sobrinus.  
 XX  
 PN JP05023188-A.  
 XX  
 PD 02-FEB-1993.  
 XX  
 PF 25-JUL-1991; 91JP-0186592.  
 XX  
 PR 25-JUL-1991; 91JP-0186592.  
 XX  
 PA (FUKU/) FUKUI I.  
 PA (KATO/) KATO K.  
 XX  
 XX WPI; 1993-079449/10.  
 DR N-PSDB; AAQ37760.  
 XX  
 PT DNA sequence glucosyl:transferase-I - comprises Streptococcus  
 PT sobrinus DNA sequence with at least one nucleotide added or  
 PT deleted  
 XX  
 PS Claim 13; Page 15; 29pp; Japanese.  
 XX  
 CC The DNA sequence from Streptococcus sobrinus strain 6715 encodes  
 CC glucosyltransferase-I (and mutants). The DNA was obtd. by treating  
 CC S. sobrinus 6715 with mutanolysin, extracting the chromosomal DNA,  
 CC partially digesting with Sau3AI and fractionating on agarose gel.  
 CC The 3-5 kbp fragment was ligated into pUC18 and E. coli JM109  
 CC transformed with it. A GT-1 expressing clone was isolated and  
 CC sequenced. The clone may be used in the development of a drug for  
 CC dental caries.  
 XX  
 SQ Sequence 1592 AA;

Query Match 54.9%; Score 62; DB 14; Length 1592;  
 Best Local Similarity 57.9%; Pred. No. 0.46;  
 Matches 11; Conservative 4; Mismatches 4; Indels 0; Gaps 0;



QY 1 AIDHLSILEAWSGNDNDYV 19  
 I :||:||||||| I :  
 Db 477 annhvsiveawsdndtpyl 495

RESULT 4  
 AAB43277  
 ID AAB43277 standard; Protein; 218 AA.  
 XX  
 AC AAB43277;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human ORFX ORF3041 polypeptide sequence SEQ ID NO:6082.  
 XX  
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnary; antiprosoratic; antiparkinsonian; nootropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200058473-A2.  
 PN  
 XX  
 PD 05-OCT-2000.  
 XX  
 XX 31-MAR-2000; 2000WO-US08621.  
 XX  
 XX 31-MAR-1999; 99US-0127607.  
 PR  
 XX 02-APR-1999; 99US-0127636.  
 PR  
 XX 05-APR-1999; 99US-0127728.  
 PR  
 XX 30-MAR-2000; 2000US-0540763.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 PA  
 XX Shimkets RA, Leach M;  
 PI  
 XX WPI; 2000-602362/57.  
 XX  
 DR N-PSDB; AAC77486.

Novel nucleic acids and peptides derived from open reading frame X,  
 useful for treating e.g. cancers, proliferative disorders,  
 neurodegenerative disorders and cardiovascular disease -  
 Claim 11; Page 5265-5266; 5507pp; English.

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 sequences have activities such as: cytostatic; hepatotropic; vulnary;  
 antiprosoratic; antiparkinsonian; nootropic; neuroprotective;  
 osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
 immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antithyroid; and antianaemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus

CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX  
 SQ Sequence 218 AA;

Query Match 42.9%; Score 48.5; DB 21; Length 218;  
 Best Local Similarity 52.6%; Pred. No. 6.1;  
 Matches 10; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

QY 3 DHLISILEAWSGNDNDYVKQ 21  
 I :||:||||||| I :  
 Db 155 drlvleswa-ndpdykr 172

RESULT 5  
 AAM25871  
 ID AAM25871 standard; Protein; 289 AA.  
 XX  
 AC AAM25871;  
 XX  
 DT 16-OCT-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:1386.  
 XX  
 KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
 KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
 KW antibacterial; endocrine; cardiant; central nervous system; virucide;  
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
 KW antiaggregant; haemostatic; vulnary; antilucer; osteopathic; eczema;  
 KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;  
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
 KW neurological disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200153455-A2.  
 PN  
 XX 26-JUL-2001.  
 XX  
 XX 22-DEC-2000; 2000WO-US35017.  
 PF  
 XX 23-DEC-1999; 99US-0471275.  
 PR  
 XX 21-JAN-2000; 2000US-0488725.  
 PR  
 XX 25-APR-2000; 2000US-0552317.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX Tang YT, Liu C, Drmanac RT;  
 PI  
 XX WPI; 2001-457603/49.  
 XX  
 DR N-PSDB; AAH99812.  
 XX  
 XX Isolated human polynucleotides encoding polypeptides, useful for the  
 XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection -  
 XX Claim 20; Page 284; 1217pp; English.

AAH99166 to AAH99904 encode the human proteins given in AAM25225 to  
 AAM25963. The proteins can have activities based on the tissues and  
 CC cells they are expressed in, such as: antiinflammatory; antirheumatic;  
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;  
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;



```

XX DR N-PSDB; AAC55231.
XX
XX PT New human nucleic acids encoding secreted proteins, useful in the
XX PT treatment, prevention or diagnosis of immune disorders (e.g. autoimmune
XX PT diseases), blood protein disorders and hyperproliferative diseases
XX PT (e.g. Gaucher's disease) -
XX PS
XX PS Claim 11; Page 456-457; 488pp; English.
XX
XX CC The polynucleotide sequences given in AAC55190 to AAC55235 encode the
XX CC human secreted proteins given in AAB32371 to AAB32484. Human secreted
XX CC proteins have activities based on the tissues and cells the genes are
XX CC expressed in. Examples of activities include: cytostatic;
XX CC immunostimulant; antiproliferative; cardiant; antiarrhythmic; antiviral;
XX CC antibacterial; antifungal; antiparasitic; neuroprotective; nontropic;
XX CC antiinflammatory; antiangiogenic; anti-HIV; and antiarteriosclerotic. The
XX CC polynucleotides and polypeptides, or their agonists and antagonists, can
XX CC be used for treating, preventing or diagnosing immune disorders (e.g.
XX CC cancer, autoimmune diseases), disorders of haematopoietic cells, blood
XX CC protein disorders (e.g. agammaglobulinaemia), hyperproliferative
XX CC diseases (e.g. Gaucher's disease), cardiovascular disorders (e.g.
XX CC congenital heart defects, pulmonary atresia, arrhythmias, ischaemia),
XX CC angiogenesis related disorders (e.g. Crohn's disease, atherosclerosis),
XX CC neurological diseases (e.g. Alzheimer's disease, Huntington's chorea),
XX CC infectious diseases (e.g. AIDS, cat-scratch disease and other bacterial,
XX CC viral, parasitic or fungal diseases). AAC55181 to AAC55189 and AAB32370
XX CC represent sequences used in the exemplification of the present
XX CC invention.
XX SQ Sequence 427 AA;
Query Match 42.9%; Score 48.5; DB 21; Length 427;
Best Local Similarity 52.6%; Pred. No. 13;
Matches 10; Conservative 4; Mismatches 4; Indels 1; Gaps 1;
Oy 3 DHLFILEAWSGNDNDYVKQ 21
Db 269 drlvtleswa-ndpdytkr 286
RESULT 8
AAB32411
ID AAB32411 standard; Protein; 436 AA.
XX AC AAB32411;
XX 16-JAN-2001 (first entry)
Human secreted protein sequence encoded by gene 14 SEQ ID NO:97.
XX KW Human; secreted protein; cytostatic; immunostimulant; antiproliferative;
XX KW cardiant; antiarrhythmic; antiviral; antibacterial; antifungal; cancer;
XX KW antiparasitic; neuroprotective; nontropic; antiinflammatory; anti-HIV;
XX KW antiangiogenic; antiarteriosclerotic; diagnosis; immune disorder; AIDS;
XX KW autoimmune disease; haematopoietic cell disorder; blood protein disorder;
XX KW agammaglobulinaemia; hyperproliferative disease; Gaucher's disease;
XX KW cardiovascular disorder; congenital heart defect; pulmonary atresia;
XX KW arrhythmia; ischaemia; angiogenesis related disorder; Crohn's disease;
XX KW atherosclerosis; neurological disease; Alzheimer's disease;
XX KW Huntington's; infectious disease; cat-scratch disease; chromosome 16.
XX OS Homo sapiens.
XX
XX PN WO200047602-A1.
XX
XX PD 17-AUG-2000.
XX
XX PF 08-FEB-2000; 2000WO-US03062.
XX
XX PR 10-FEB-1999; 99US-0119468.

```

---

```

XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM, Ebner R, Young PE, Ni J, Soppet DR, Moore PA;
XX PI Shi Y, Lafleur DW, Olsen HS, Florence KA, Komatsoulis G;
XX DR WPI; 2000-543578/49.
XX DR N-PSDB; AAC55230.
XX
XX PT New human nucleic acids encoding secreted proteins, useful in the
XX PT treatment, prevention or diagnosis of immune disorders (e.g. autoimmune
XX PT diseases), blood protein disorders and hyperproliferative diseases
XX PT (e.g. Gaucher's disease) -
XX PS
XX PS Claim 11; Page 454-455; 488pp; English.
XX
XX CC The polynucleotide sequences given in AAC55190 to AAC55235 encode the
XX CC human secreted proteins given in AAB32371 to AAB32484. Human secreted
XX CC proteins have activities based on the tissues and cells the genes are
XX CC expressed in. Examples of activities include: cytostatic;
XX CC immunostimulant; antiproliferative; cardiant; antiarrhythmic; antiviral;
XX CC antibacterial; antifungal; antiparasitic; neuroprotective; nontropic;
XX CC antiinflammatory; antiangiogenic; anti-HIV; and antiarteriosclerotic. The
XX CC polynucleotides and polypeptides, or their agonists and antagonists, can
XX CC be used for treating, preventing or diagnosing immune disorders (e.g.
XX CC cancer, autoimmune diseases), disorders of haematopoietic cells, blood
XX CC protein disorders (e.g. agammaglobulinaemia), hyperproliferative
XX CC diseases (e.g. Gaucher's disease), cardiovascular disorders (e.g.
XX CC congenital heart defects, pulmonary atresia, arrhythmias, ischaemia),
XX CC angiogenesis related disorders (e.g. Crohn's disease, atherosclerosis),
XX CC neurological diseases (e.g. Alzheimer's disease, Huntington's chorea),
XX CC infectious diseases (e.g. AIDS, cat-scratch disease and other bacterial,
XX CC viral, parasitic or fungal diseases). AAC55181 to AAC55189 and AAB32370
XX CC represent sequences used in the exemplification of the present
XX CC invention.
XX SQ Sequence 436 AA;
Query Match 42.9%; Score 48.5; DB 21; Length 436;
Best Local Similarity 52.6%; Pred. No. 14;
Matches 10; Conservative 4; Mismatches 4; Indels 1; Gaps 1;
Oy 3 DHLFILEAWSGNDNDYVKQ 21
Db 270 drlvtleswa-ndpdytkr 287
RESULT 9
AAB32384
ID AAB32384 standard; Protein; 576 AA.
XX AC AAB32384;
XX 16-JAN-2001 (first entry)
Human secreted protein sequence encoded by gene 14 SEQ ID NO:70.
XX KW Human; secreted protein; cytostatic; immunostimulant; antiproliferative;
XX KW cardiant; antiarrhythmic; antiviral; antibacterial; antifungal; cancer;
XX KW antiparasitic; neuroprotective; nontropic; antiinflammatory; anti-HIV;
XX KW antiangiogenic; antiarteriosclerotic; diagnosis; immune disorder; AIDS;
XX KW autoimmune disease; haematopoietic cell disorder; blood protein disorder;
XX KW agammaglobulinaemia; hyperproliferative disease; Gaucher's disease;
XX KW cardiovascular disorder; congenital heart defect; pulmonary atresia;
XX KW arrhythmia; ischaemia; angiogenesis related disorder; Crohn's disease;
XX KW atherosclerosis; neurological disease; Alzheimer's disease;
XX KW Huntington's; infectious disease; cat-scratch disease; chromosome 16.
XX OS Homo sapiens.
XX
XX PN WO200047602-A1.
XX
XX PD 17-AUG-2000.
XX
XX PF 08-FEB-2000; 2000WO-US03062.
XX
XX PR 10-FEB-1999; 99US-0119468.

```

PD 17-AUG-2000.  
 XX  
 PF 08-FEB-2000; 2000WO-US03062.  
 XX  
 PR 10-FEB-1999; 99US-0119468.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM, Ebner R, Young PE, Ni J, Soppet DR, Moore PA;  
 PI Shi Y, Lefleur DW, Olsen HS, Florence KA, Komatsoulis G;  
 XX  
 DR WPI; 2000-543578/49.  
 DR N-PSDB; AAC55203.  
 XX  
 PT New human nucleic acids encoding secreted proteins, useful in the  
 PT treatment, prevention or diagnosis of immune disorders (e.g. autoimmune  
 PT diseases), blood protein disorders and hyperproliferative diseases  
 (e.g. Gaucher's disease) -  
 Claim 11; Page 432-434; 488pp; English.  
 XX  
 CC The polynucleotide sequences given in AAC55190 to AAC55235 encode the  
 CC human secreted proteins given in AAB32371 to AAB32484. Human secreted  
 CC proteins have activities based on the tissues and cells the genes are  
 CC expressed in. Examples of activities include: cytostatic;  
 CC immunostimulant; antiproliferative; cardiant; antiarrhythmic; antiviral;  
 CC antibacterial; antifungal; antiparasitic; neuroprotective; nootropic;  
 CC antiinflammatory; angiogenic; anti-HIV; and antiarteriosclerotic. The  
 CC polynucleotides and polypeptides, or their agonists and antagonists, can  
 CC be used for treating, preventing or diagnosing immune disorders (e.g.  
 CC cancer, autoimmune diseases), disorders of haematopoietic cells, blood  
 CC protein disorders (e.g. agammaglobulinaemia), hyperproliferative  
 CC diseases (e.g. Gaucher's disease), cardiovascular disorders (e.g.  
 CC congenital heart defects, pulmonary atresia, arrhythmias, ischaemia),  
 CC angiogenesis related disorders (e.g. Crohn's disease, atherosclerosis),  
 CC neurological diseases (e.g. Alzheimer's disease, Huntington's chorea),  
 CC infectious diseases (e.g. AIDS, cat-scratch disease and other bacterial,  
 CC viral, parasitic or fungal diseases). AAC55181 to AAC55189 and AAB32370  
 CC represent sequences used in the exemplification of the present  
 CC invention.  
 XX Sequence 576 AA;  
 SQ  
 Query Match 42.9%; Score 48.5; DB 21; Length 576;  
 Best Local Similarity 52.6%; Pred. No. 19;  
 Matches 10; Conservative 4; Mismatches 4; Indels 1; Gaps 1;  
 Db 3 DHLSTLEAWSGNDNDYVKQ 21  
 | | | | | | | | | |  
 270 drlvteswa-ndpdyklr 287  
 RESULT 10  
 AAB94297  
 ID AAB94297 standard; Protein; 576 AA.  
 XX  
 AC AAB94297;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:14751.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX

29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 8; SEQ ID 14751; 2537pp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesising 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesising polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX Sequence 576 AA;  
 SQ  
 Query Match 42.9%; Score 48.5; DB 22; Length 576;  
 Best Local Similarity 52.6%; Pred. No. 19;  
 Matches 10; Conservative 4; Mismatches 4; Indels 1; Gaps 1;  
 QY 3 DHLSTLEAWSGNDNDYVKQ 21  
 | | | | | | | | | |  
 Db 270 drlvteswa-ndpdyklr 287  
 RESULT 11  
 AAY34618  
 ID AAY34618 standard; Protein; 661 AA.  
 XX  
 AC AAY34618;  
 XX  
 DT 13-SEP-1999 (first entry)  
 XX  
 DE Chlamydia pneumoniae surface exposed polypeptide.  
 XX  
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope.  
 XX  
 OS Chlamydia pneumoniae.  
 XX  
 PN WO9927105-A2.  
 XX

PD 03-JUN-1999.  
 XX 20-NOV-1998; 98WO-IB01890.  
 XX 04-NOV-1998; 98US-0107078.  
 PR 21-NOV-1997; 97FR-0014673.  
 XX (GEST ) GENSET.  
 FA Griffais R;  
 XX WPI; 1999-357842/30.  
 XX Genome sequence of Chlamydia pneumoniae  
 PT Page 638-639; Disclosure; 1912pp; English.  
 XX  
 XX AAY34584-Y35879 represent the proteins encoded by all the open reading  
 CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.  
 CC C. pneumoniae causes respiratory disease such as pneumonia and  
 CC bronchitis and is thought to be a contributing factor in heart  
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading  
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in  
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
 CC nucleotide sequences can also be used as immunogenic compositions,  
 CC especially where the vector directs the expression of a neutralising  
 CC epitope of C. pneumoniae.  
 XX Sequence 661 AA;  
 SQ

Query Match 42.5%; Score 48; DB 20; Length 661;  
 Best Local Similarity 57.1%; Pred. No. 27;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 Qy 3 DHLSTLEAWSGNDN 16  
 Db 572 dhmlrlewigndn 585  
 RESULT 12  
 ID AAB82125 standard; Protein; 722 AA.  
 XX  
 AC AAB82125;  
 XX  
 DT 13-JUL-2001 (first entry)  
 XX Alpha-1,3-multi-branch dextran hydrolase.  
 XX Alpha-1,3-multi-branch dextran hydrolase; enzyme.  
 XX Spingobacterium sp. V-54.  
 XX JP2001054382-A.  
 XX 27-FEB-2001.  
 XX  
 PF 16-AUG-1999; 99JP-0229541.  
 XX  
 PR 16-AUG-1999; 99JP-0229541.  
 XX  
 PA (NODA ) 2H NODA SANGYO KAGAKU KENKYUSHO.  
 XX  
 DR WPI; 2001-294617/31.  
 DR N-PSDB; AAF86585.  
 XX  
 PT Alpha-1,3-multi-branch dextran hydrolase gene, for the recombinant  
 preparation of alpha-1,3-multi-branch dextran hydrolase  
 XX  
 PS Claim 1; Page 7-8; 10pp; Japanese.  
 XX

CC The present invention relates to alpha-1,3-multi-branch dextran hydrolase  
 CC gene (see AAF86585), derived from Spingobacterium sp. V-54. The gene  
 CC sequence can be used for the recombinant preparation of  
 CC alpha-1,3-multi-branch dextran hydrolase (the present sequence).  
 XX  
 SQ Sequence 722 AA;  
 Query Match 41.6%; Score 47; DB 22; Length 722;  
 Best Local Similarity 72.7%; Pred. No. 43;  
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Qy 9 BAWSGNDNDYV 19  
 Db 397 edwfgndkdyv 407  
 RESULT 13  
 AAG28898  
 ID AAG28898 standard; Protein; 730 AA.  
 XX  
 AC AAG28898;  
 XX  
 DT 17-OCT-2000 (first entry)  
 XX Arabidopsis thaliana protein fragment SEQ ID NO: 34287.  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 34287.  
 XX Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX Arabidopsis thaliana.  
 OS Arabidopsis thaliana.  
 XX  
 PN EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 25-FEB-2000; 2000EP-0301439.  
 XX  
 PR 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 18-MAY-1999; 99US-0134370.  
 PR 19-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR

PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137508.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140693.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142053.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 09-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150586.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 41.6%; Score 47; DB 21; Length 730;  
Best Local Similarity 66.7%; Pred. No. 44;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Oy 5 LSILEAMSGNDN 16  
Db 477 vsilaawtgnds 488

## RESULT 14

AAG28897  
ID AAG28897 standard; Protein; 748 AA.

XX

AC AAG28897;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 34286.

DE Protein identification; signal transduction pathway; metabolic pathway;  
XX hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.

XX Arabidopsis thaliana.

OS EPI033405-A2.

PN 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.

PR 10-AUG-1999; 99US-01481171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 41.6%; Score 47; DB 21; Length 748;  
Best Local Similarity 66.7%; Pred. No. 45;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
QY 5 LSILEAMSGNDN 16  
:|||||:|||||

Db 495 vsilaawtgnds 506  
RESULT 15  
AAG28896  
ID AAG28896 standard; Protein; 769 AA.  
XX AC AAG28896;  
XX DT 17-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 34285.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EPI033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127452.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.



PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144088.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.

PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 41.6%; Score 47; DB 21; Length 769;  
Best Local Similarity 66.7%; Pred No. 47;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
QY 5 LSILEAWGNDN 16  
Db :||| |||||  
516 vsilaawtgnds 527

Search completed: March 27, 2002, 13:57:57  
Job time: 525 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:27:02 ; Search time 53.4 Seconds  
(without alignments)  
14.419 Million cell updates/sec

Title: US-09-290-049a-11  
Perfect score: 116  
Sequence: 1 AINHL5ILEAWSNDPQYNKD 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

1 number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query %	Length	ID	Description
1	116	100.0	1462	1	GTFFD_STRMU
2	81	69.8	1365	1	GTFFS_STRDO
3	75	64.7	1476	1	GTFFB_STRMU
4	74	63.8	1592	1	GTFFD_STRDO
5	74	63.8	1597	1	GTFL_STRDO
6	66	56.9	1375	1	GTFC_STRMU
7	51	44.0	429	1	HPAG_SALDU
8	47	40.5	418	1	VS12_REOVL
9	46	39.7	418	1	VS12_REOVL
10	45.5	39.2	560	1	PMG2_ARATH
11	45	38.8	657	1	CSPL_CORGL
12	45	38.8	2358	1	MOKD_SCHPO
13	44.5	38.4	241	1	HEBL_HAEIN
14	44.5	38.4	241	1	HEB2_HAEIN
15	44	37.9	837	1	ROD1_YEAST
16	43.5	37.5	825	1	ILAR_HUMAN
17	43	37.1	267	1	COML_NEIGO
18	43	37.1	418	1	VS12_REOVL
19	43	37.1	443	1	AMZA_ORYSA
20	43	37.1	445	1	AMC2_ORYSA
21	43	37.1	480	1	DISA_TRIGA
22	42.5	36.6	451	1	MYBH_DICDI
23	42	36.2	405	1	HPCE_ECOLI
24	42	36.2	429	1	HPAG_ECOLI
25	42	36.2	447	1	TRN1_RAT
26	42	36.2	461	1	TRN1_RAT
27	42	36.2	989	1	RPOC_LEUME
28	42	36.2	1097	1	LIFR_HUMAN
29	41.5	35.8	1616	1	YABD_SCHPO
30	41	35.3	305	1	SGAU_MYCPN
31	41	35.3	576	1	CBPS_YEAST
32	41	35.3	591	1	CALX_MOUSE
33	41	35.3	591	1	CALX_RAT

34	41	35.3	592	1	CALX_HUMAN
35	41	35.3	622	1	VAIL_TREPA
36	41	35.3	672	1	VAIL_HUMAN
37	41	35.3	674	1	CORA_BACTJ
38	41	35.3	989	1	RPOC_LEUPS
39	41	35.3	1308	1	YTX2_XENLA
40	41	35.3	2410	1	MOK1_SCHPO
41	40.5	34.9	525	1	RRPO_IBDV5
42	40.5	34.9	598	1	NBL4_HUMAN
43	40.5	34.9	878	1	RRPO_IBDVA
44	40.5	34.9	1055	1	CC28_SCHPO
45	40.5	34.9	1131	1	YPO4_CAEEL

## ALIGNMENTS

RESULT 1  
GTFFD\_STRMU STANDARD: PRT: 1462 AA.  
AC P49331: O69383; O69386; O69389; O69392; O69398;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)  
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
CN GTFD.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-GS-5;  
RX MEDLINE=91100958; PubMed=2148600;  
RA Honda O., Kato C., Kuramitsu H.K.;  
RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding  
the glucosyltransferase-S enzyme";  
RL J. Gen. Microbiol. 136:2099-2105(1990).  
RN [2]  
SEQUENCE FROM N.A.  
RC STRAIN-MT4239, MT4245, MT4251, MT4467, AND MT8148;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
Kimura S., Hamada S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of  
Streptococcus mutans";  
RL FEMS Microbiol. Lett. 161:331-336(1998).  
CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE  
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -!- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =  
D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
CC -!- SUBCELLULAR LOCATION: SECRETED.  
CC -!- DISEASE: DENTAL CARIES.  
CC -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
FORMS OF GLUCANS.  
CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
BINDING PROTEIN FROM S. MUTANS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
-----  
DR EMBL; M29296; AAA26895.1; -;  
DR EMBL; D88653; BAA26103.1; -;

DR	EMBL; D88656;	BAA26107.1;	-
DR	EMBL; D88659;	BAA26111.1;	-
DR	EMBL; D88662;	BAA26115.1;	-
DR	EMBL; D89979;	BAA26121.1;	-
DR	InterPro; IPR002479;	CW_binding.	
DR	InterPro; IPR003318;	Glyco_hydro_70.	
DR	Pfam; PF01473;	Cw_binding_1;	I.
DR	Pfam; PF02324;	Glyco_hydro_70;	1.
KW	transferase; Glycosyltransferase;	Signal; Repeat; Dental caries.	
KW	SIGNAL	? POTENTIAL.	
FT	CHAIN	? 1462	
FT	DOMAIN	1232 1433	GLUCOSYLTRANSFERASE-S.
FT	REPEAT	1232 1295	3 X 63 AA APPROXIMATE TANDEM REPEATS.
FT	REPEAT	1296 1359	1.
FT	REPEAT	1360 1423	2.
FT	VARIANT	58 58	3.
FT	VARIANT	68 68	K -> E (IN STRAIN MT4467).
FT	VARIANT	81 81	A -> S (IN STRAINS MT4239 AND MT4245).
FT	VARIANT	113 113	A -> T (IN STRAINS MT4251 AND MT8148).
FT	VARIANT	122 132	T -> I (IN STRAINS MT4239 AND MT4245).
FT	VARIANT	132 132	A -> V (IN STRAINS MT4239, MT4245 AND MT8148).
FT	VARIANT	132 132	S -> A (IN STRAINS MT4239, MT4245, MT4251 AND MT8148).
FT	VARIANT	135 135	A -> V (IN STRAIN MT4245).
FT	VARIANT	202 202	V -> L (IN STRAIN MT4239).
FT	VARIANT	255 255	D -> N (IN STRAIN MT8148).
FT	VARIANT	275 275	E -> D (IN STRAINS MT4239, MT4245 AND MT4251).
FT	VARIANT	288 288	D -> N (IN STRAINS MT4239, MT4245 AND MT4251).
FT	VARIANT	301 301	Q -> H (IN STRAIN MT4245).
FT	VARIANT	313 313	D -> N (IN STRAINS MT4239 AND MT4251).
FT	VARIANT	317 317	E -> K (IN STRAIN MT4239).
FT	VARIANT	328 328	V -> F (IN STRAIN MT4239).
FT	VARIANT	350 350	F -> L (IN STRAINS MT4239, MT4251 AND MT4467).
FT	VARIANT	628 633	KKKYTQ -> EKVEYTL (IN STRAIN MT4251).
FT	VARIANT	688 688	ADQSGE -> S (IN STRAIN MT4239).
FT	VARIANT	726 732	TDQGSS -> ADKGNDS (IN STRAIN MT4251).
FT	VARIANT	726 730	TDQSIS -> ADKGN (IN STRAINS MT4239 AND MT4245).
FT	VARIANT	964 964	D -> Y (IN STRAIN MT4251).
FT	VARIANT	1019 1019	E -> K (IN STRAINS MT4245 AND MT4251).
FT	VARIANT	1059 1060	LG -> IR (IN STRAIN MT4251).
FT	VARIANT	1060 1060	G -> R (IN STRAIN MT4245).
FT	VARIANT	1060 1060	G -> H (IN STRAIN MT4239).
FT	VARIANT	1142 1142	Q -> H (IN STRAINS MT4239, MT4245, MT4251, MT4467 AND MT8148).
FT	VARIANT	1198 1198	S -> N (IN STRAIN MT4239).
FT	VARIANT	1220 1220	S -> C (IN STRAINS MT4251 AND MT4467).
FT	VARIANT	1280 1280	F -> L (IN STRAIN MT4467).
FT	VARIANT	1282 1282	Q -> P (IN STRAIN MT4245).
FT	VARIANT	1290 1290	K -> T (IN STRAIN MT4245).
FT	VARIANT	1311 1311	N -> D (IN STRAIN MT4245).
FT	VARIANT	1403 1403	D -> G (IN STRAINS MT4239, MT4245, MT4251 AND MT8148).
FT	VARIANT	1425 1425	R -> G (IN STRAINS MT4239, MT4245, MT4251, MT4467 AND MT8148).
FT	VARIANT	1449 1449	R -> K (IN STRAIN MT4467).
FT	CONFLICT	1428 1462	RYDKNSGVNMYNKVVTLANRRIGIDRWGIARYYY -> WVR (IN REF. 1).
SQ	SEQUENCE	1462 AA:	MW: 5C6541FDCBDF00 CRC64:

```

RESULT 2
GTFS_STRDO STANDARD; PRT; 1365 AA.
ID GTFS_STRDO STANDARD; PRT; 1365 AA.
P29336;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
DN GTFS.
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
ON NCBI_TaxID=1317;
OX [1]
RC SEQUENCE FROM N.A.
RC STRAIN=MFE28;
RX MEDLINE=90316665; PubMed=2142479;
RA Gilmore K.S., Russell R.R., Ferretti J.J.;
RT "Analysis of the Streptococcus downei gtfS gene, which specifies a
RT glucosyltransferase that synthesizes soluble glucans.";
RL Infect. Immun. 58:2452-2458(1990).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).
CC -1- ENZYME REGULATION: GLUCAN SYNTHESIS BY GTF-S IS INDEPENDENT OF
CC PRIMER GLUCAN UNLIKE GTF-1.
CC -1- DISEASE: DENTAL CARIES.
CC -1- MISCELLANEOUS: GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA
CC 1,6-GLUCOSE).
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S.MUTANS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; M30943; AAA26898.1; -.
DR PIR; A41483;
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding.1; 10.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1 36 OR 37 (POTENTIAL).
FT CHAIN 37 1365 GLUCOSYLTRANSFERASE-S.
FT DOMAIN 37 1050 CATALYTIC (APPROXIMATE).
FT FT DOMAIN 1083 1365 GLUCAN-BINDING (APPROXIMATE).
FT FT DOMAIN 1083 1365 4-5 X TANDEM REPEATS.
FT REPEAT 1083 1131 1.
FT REPEAT 1150 1199 2.
FT REPEAT 1225 1274 3.
FT REPEAT 1289 1339 4.
FT REPEAT 1353 1365 5 (INCOMPLETE).
FO SEQUENCE 1365 AA; 151590 MW; 167296B5A2E8C476 CRC64:

```

RESULT 3  
GTFB\_STRMU STANDARD; PRT: 1476 AA.  
AC P08987; 069381; 069384; 069387; 069390; 069396;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)  
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
GN GTFB.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GS-5;  
RC MEDLINE=87308013; PubMed=3040685;  
RX Shiroza T., Ueda S., Kuramitsu H.K.;  
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";  
J. Bacteriol. 169:4263-4270(1987).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MT4239, MT4245, MT4251, MT4467, AND MT8148;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
RA Kimura S., Hamada S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of  
Streptococcus mutans.";  
RL FEMS Microbiol. Lett. 161:331-336(1998).  
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE  
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) =  
D-FRUCTULOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).  
CC -1- SUBCELLULAR LOCATION: SECRETED.  
CC -1- DISEASE: DENTAL CARIES.  
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
FORMS OF GLUCANS.  
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
BINDING PROTEIN FROM S.MUTANS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; M17361; AAA88588.1; -  
DR EMBL; D88651; BAA26101.1; -  
DR EMBL; D88654; BAA26105.1; -  
DR EMBL; D88657; BAA26109.1; -  
DR EMBL; D88660; BAA26113.1; -  
DR EMBL; D89977; BAA26119.1; -  
DR PIR; B33135; B33135.  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 13.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.  
FT SIGNAL 1 34 POTENTIAL.  
FT CHAIN 35 1476 GLUCOSYLTRANSFERASE-I.  
FT DOMAIN 35 1051 CATALYTIC (APPROXIMATE).  
FT DOMAIN 1097 1476 GLUCAN-BINDING (APPROXIMATE).  
FT REPEAT 1097 1130 A REPEAT.  
FT DOMAIN 1161 1470 5 X TANDEM REPEATS.  
FT REPEAT 1161 1210 1.  
FT REPEAT 1225 1275 2.

FT REPEAT 1290 1340 3.  
FT REPEAT 1355 1405 4.  
FT REPEAT 1420 1470 5.  
FT VARIANT 62 62 S -> T (IN STRAIN MT4239).  
FT VARIANT 65 65 MT4251, MT4467 AND MT8148).  
FT VARIANT 68 68 A -> V (IN STRAIN MT4239).  
FT VARIANT 78 78 O -> P (IN STRAIN MT4251).  
FT VARIANT 86 86 S -> I (IN STRAIN MT4239).  
FT VARIANT 89 89 S -> F (IN STRAIN MT4251).  
FT VARIANT 168 168 K -> N (IN STRAIN MT4251).  
FT VARIANT 276 276 D -> S (IN STRAINS MT4239, MT4245 AND MT4251).  
FT VARIANT 399 399 N -> R (IN STRAIN MT4239).  
FT VARIANT 474 474 I -> T (IN STRAIN MT4239).  
FT VARIANT 512 512 K -> R (IN STRAIN MT8148).  
FT VARIANT 519 519 F -> Y (IN STRAIN MT8148).  
FT VARIANT 701 701 T -> I (IN STRAIN MT8148).  
FT VARIANT 708 708 A -> V (IN STRAIN MT8148).  
FT VARIANT 938 938 F -> L (IN STRAIN MT8148).  
FT VARIANT 952 957 YGTPVA -> FGKPE (IN STRAINS MT4245, MT4251 AND MT8148).  
FT VARIANT 963 964 NT -> SV (IN STRAINS MT4245, MT4251 AND MT8148).  
FT VARIANT 968 970 VDG -> ADS (IN STRAINS MT4245, MT4251 AND MT8148).  
FT VARIANT 1086 1086 A -> T (IN STRAIN MT4239).  
FT VARIANT 1158 1158 S -> N (IN STRAIN MT4239).  
FT VARIANT 1163 1163 H -> Y (IN STRAIN MT4251).  
FT VARIANT 1168 1168 E -> K (IN STRAIN MT8148).  
FT VARIANT 1182 1182 Y -> C (IN STRAIN MT8148).  
FT VARIANT 1234 1234 A -> P (IN STRAIN MT4239).  
FT VARIANT 1263 1263 H -> P (IN STRAIN MT8148).  
FT VARIANT 1263 1263 H -> R (IN STRAINS MT4239, MT4245 AND MT4251).  
FT VARIANT 1264 1264 H -> Y (IN STRAINS MT4245 AND MT4251).  
FT VARIANT 1272 1272 G -> S (IN STRAINS MT4245 AND MT4251).  
FT VARIANT 1329 1329 Y -> H (IN STRAINS MT4239, MT4245, MT4251 AND MT8148).  
FT VARIANT 1394 1394 H -> Y (IN STRAINS MT4245 AND MT4251).  
FT VARIANT 1402 1402 G -> S (IN STRAINS MT4245 AND MT4251).  
FT VARIANT 1459 1459 Y -> H (IN STRAIN MT4467).  
FT CONFLICT 570 570 R -> A (IN REF. 1).  
FT CONFLICT 800 817 ADDVRVAASTAPSTGK -> LKMFALRLARPHQOMA (IN REF. 1).  
FT CONFLICT 1310 1310 H -> L (IN REF. 1).  
SQ SEQUENCE 1476 AA; 165685 MW; 3479B62B07694D98 CRC64;  
Query Match 64.7%; Score 75; DB 1; Length 1476;  
Best Local Similarity 71.4%; Pred. No. 0.0017;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
QY 1 AINHLISLEAWSNDPQYKND 21  
I : :::::::::::::: I  
Db 481 ANDHLISLEAWSNDTPYLHD 501  
RESULT 4  
GTF2\_STRDO STANDARD; PRT; 1592 AA.  
ID GTF2\_STRDO  
AC P27470;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)  
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
OS Streptococcus downei (Streptococcus sobrinus).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1317;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=6715;  
RX MEDLINE=91123227; PubMed=1704006;  
RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,  
RA Kagawa H.;  
RT "Peptide sequences for sucrose splitting and glucan binding within  
RT streptococcus sobrinus glucosyltransferase (water-insoluble glucan  
RT synthetase).";  
RL J. Bacteriol. 173:989-996 (1991).  
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =  
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
CC -1- SUBCELLULAR LOCATION: SECRETED.  
CC -1- DISEASE: DENTAL CARIES.  
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
CC FORMS OF GLUCANS.  
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
CC BINDING PROTEIN FROM S. MUTANS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; D90213; BAA14241.1; -;  
CC PIR; A38175; A38175.  
CC HSP; P00695; 2HEE.  
CC InterPro; IPR002479; CW\_binding.  
CC InterPro; IPR003318; Glyco\_hydro\_70.  
CC Pfam; PF01473; CW\_binding\_1; 16.  
CC Pfam; PF02324; Glyco\_hydro\_70; 1.  
CC Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.  
CC KW SIGNAL 1 38 POTENTIAL.  
CC FT CHAIN 39 1592 GLUCOSYLTRANSFERASE-I.  
CC FT DOMAIN 39 1044 CATALYTIC (APPROXIMATE).  
CC FT DOMAIN 1093 1592 GLUCAN-BINDING (APPROXIMATE).  
CC FT DOMAIN 1093 1592 6.5 X TANDEM REPEATS.  
CC FT REPEAT 1093 1142 1.  
CC FT REPEAT 1158 1207 2.  
CC FT REPEAT 1222 1272 3.  
CC FT REPEAT 1287 1337 4.  
CC FT REPEAT 1402 1451 5.  
CC FT REPEAT 1514 1563 6.  
CC FT REPEAT 1577 1592 7 (INCOMPLETE).  
CC SQ SEQUENCE 1592 AA; 176167 MW; BCOA66D079351ECF CRC64;

Query Match 63.8%; Score 74; DB 1; Length 1592;  
Best Local Similarity 66.7%; Pred. No. 0.0026;  
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 AINHLSILEAWSNDNDPOYNKD 21  
DB 477 ANNHVSIVEAWSNDNDTPYLHD 497  
RESULT 5  
GTF1\_STRDO STANDARD; PRT; 1597 AA.  
ID GTF1\_STRDO  
AC P11001;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)  
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
GN GTFI.

OS Streptococcus downei (Streptococcus sobrinus).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_taxid=1317;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=MFE28;  
RX MEDLINE=87308014; PubMed=3040686;  
RA Ferretti J.J., Gilpin M.L., Russell R.R.B.;  
RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus  
RT sobrinus MFE28.";  
RL J. Bacteriol. 169:4271-4278 (1987).  
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =  
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
CC -1- SUBCELLULAR LOCATION: SECRETED.  
CC -1- DISEASE: DENTAL CARIES.  
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES BOTH  
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
CC FORMS OF GLUCANS.  
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
CC BINDING PROTEIN FROM S. MUTANS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M17391; AAC63063.1; -;  
CC InterPro; IPR002479; CW\_binding.  
CC InterPro; IPR003318; Glyco\_hydro\_70.  
CC Pfam; PF01473; CW\_binding\_1; 19.  
CC Pfam; PF02324; Glyco\_hydro\_70; 1.  
CC Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.  
CC KW SIGNAL 1 38 POTENTIAL.  
CC FT CHAIN 39 1597 GLUCOSYLTRANSFERASE-I.  
CC FT DOMAIN 39 1050 CATALYTIC (APPROXIMATE).  
CC FT DOMAIN 1099 1597 GLUCAN-BINDING (APPROXIMATE).  
CC FT DOMAIN 1099 1597 1.25 A, 2 B, AND 5 AC REPEATS.  
CC FT REPEAT 1099 1132 A REPEAT.  
CC FT REPEAT 1163 1213 AC REPEAT.  
CC FT REPEAT 1227 1277 AC REPEAT.  
CC FT REPEAT 1292 1342 AC REPEAT.  
CC FT REPEAT 1352 1399 B REPEAT.  
CC FT REPEAT 1406 1455 AC REPEAT.  
CC FT REPEAT 1465 1512 B REPEAT.  
CC FT REPEAT 1519 1568 AC REPEAT.  
CC FT REPEAT 1582 1597 A REPEAT (INCOMPLETE).  
CC SQ SEQUENCE 1597 AA; 177080 MW; B9E86A200868798E CRC64;

Query Match 63.8%; Score 74; DB 1; Length 1597;  
Best Local Similarity 66.7%; Pred. No. 0.0026;  
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 AINHLSILEAWSNDNDPOYNKD 21  
DB 483 ANNHVSIVEAWSNDNDTPYLHD 503  
RESULT 6  
GTF1\_STRMU STANDARD; PRT; 1375 AA.  
ID GTF1\_STRMU  
AC P13470; P05427;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)

D6 507 ANDHLSILEAWSNDTPYLHD 527

```
RESULT      7  
HPAG_SALDU STANDARD; PRT;    429 AA.  
ID HPAG-SALDU STANDARD; PRT;    429 AA.  
AC QRPUS;  
DT DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE 4-HYDROXYPHENYLACETATE DEGRADATION BIFUNCTIONAL  
ISOMERASE/DSCARBOXYLASE [INCLUDES: 2-HYDROXIHEPTA-2,4-DIENE-1,7-DIOATE  
ISOMERASE (EC 5.3.3.-) (HHDD ISOMERASE); 5-CARBOXYMETHYL-2-OXO-  
HEX-3-ENE-1,7-DIOATE DECARBOXYLASE (EC 4.1.1.-) (OPET DECARBOXYLASE)].  
GN HPAE.  
OS Salmonella dublin.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
NCBI_TaxID=98360;  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=2229;  
RA Galyov E.E., Wood M.W., Hedges S.;  
RL "Characterization of the hpa genetic locus from Salmonella dublin.";  
RT Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: DECARBOXYLATES OPBT (5-OXO-PENT-3-ENE-1,2,5-  
TRICARBOKLYLIC ACID) INTO HHDD (2-HYDROXY-HEPT-2,4-DIENE-1,7-  
DIOATE) AND ISOMERIZES IT TO OHED (2-OXO-HEPT-3-ENE-1,7-DIOATE).  
CC -1- PATHWAY: 4-HYDROXYPHENYLACETATE DEGRADATION.  
CC -1- SIMILARITY: BELONGS TO THE FAH / HPEC FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation,  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to licensee@isb-sib.ch).  
-----  
DR ENBL: AF144422; AAD53501.1; -  
DR Interpro: IPR002529; FAA_Hydrolase.  
DR Pfam: PF01557; FAA_hydrolyase; 2.  
KW Aromatic hydrocarbons catabolism; Lyase; Isomerase; Repeat;  
MW Multifunctional enzyme.  
FT REPEAT     1   215 APPROXIMATE.  
FT REPEAT     216   429 APPROXIMATE.  
SQ SEQUENCE    429 AA; 47185 MW; D32723999AC4084E CRC64;
```

Query Match 44.0%; Score 51; DB 1; Length 429;  
Best Local Similarity 47.6%; Pred No. 1.9;  
Matches 10; Conservative 4; Mismatches 5; Indels 2; Gaps 1;

Qy 1 AI NHLSILEAWSD--NPDQYN 19  
| | | | : | : | : | : | : |  
Db 9 ALNHRSQLDAWAEFSQQPPN 29

```
RESULT      8  
VS12_REOVLL STANDARD; PRT;    418 AA.  
ID VS12_REOVLL STANDARD; PRT;    418 AA.  
AC PI1314;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 01-OCT-1994 (Rel. 30, Last annotation update)  
DE SIGMA 2 PROTEIN (CORE PROTEIN).  
GN S2.  
OS Reovirus (type 1 / strain Lang).  
OC Viruses; dsRNA viruses; Reoviridae; Orthoreovirus.  
ON NCBI_TaxID=10884;  
RX MEDLINE=G2015462; PubMed=1920614;  
SEQUENCE FROM N.A.
```





DE PS1 PROTEIN PRECURSOR.  
GN CSPI.  
OS Corynebacterium glutamicum (Brevibacterium flavum).  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
OC Corynebacterium.  
OX NCBI\_TaxID=1718;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 17965 / Melassecola;  
RX MEDLINE=93023863; PubMed=1406274;  
RA Joffe G., Mathieu L., Hahn V., Bayan N., Duchiron F., Renaud M.,  
RA Scher E., Leblon G.  
RT "Cloning and nucleotide sequence of the cspl gene encoding PS1, one  
RT of the two major secreted proteins of Corynebacterium glutamicum: the  
RT deduced N-terminal region of PS1 is similar to the Mycobacterium  
RT antigen 85 complex."  
RL MOL. Microbiol. 6:2349-2362(1992).  
CC -1- FUNCTION: ONE OF THE TWO MAJOR SECRETED PROTEINS.  
CC -1- SUBCELLULAR LOCATION: SECRETED.  
CC -1- SIMILARITY: THE N-TERMINAL IS VERY SIMILAR TO THE COMPLETE  
CC SEQUENCES OF THE MYCOBACTERIAL ANTIGENS 85-A, 85-B & 85-C.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X66078; CAA46877.1; .  
DR PIR; S25184; S25184.  
DR InterPro; IPR000379; Est\_lip\_thioest\_actsite.  
DR InterPro; IPR000801; Esterase\_put.  
DR Pfam; PF00756; Esterase; 1.  
KW Signal.  
FT SIGNAL 1 43 PS1 PROTEIN.  
FT CHAIN 44 657  
FT MOD\_RES 44 44 PYRROLIDONE CARBOXYLIC ACID (PROBABLE).  
SQ SEQUENCE 657 AA; 70918 MW; 5164B87626AC098F CRC64;  
Query Match 38.8%; Score 45; DB 1; Length 657;  
Best Local Similarity 53.8%; Pred. No. 27;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 9 EAWSDNDPQYKND 21  
| | : | | | : | |  
286 ERWQENDPKSNVD 298

RESULT 12  
MOKD\_SCHPO STANDARD; PRT; 2358 AA.  
AC Q9Y719; O94638;  
DT 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE CELL WALL ALPHA-1,3-GLUCAN SYNTHASE MOK13 (EC 2.4.1.183).  
GN MOK13 OR SPBC16D10.05.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomyces.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RA Katayama S., Dai H., Arellano M., Perez P., Toda T.;  
RT "Fission yeast alpha-glucan synthase MOK1 localizes closely with actin  
RT and play a role essential for cell morphogenesis and protein kinase C  
RT function.";

RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RA Wood V., Rajandream M.A., Barrell B.G., Brown D., Churcher C.M.;  
RA Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RL -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + (ALPHA-D-GLUCOSYL-(1,3))(N) =  
CC UDP + (ALPHA-D-GLUCOSYL-(1,3))(N+1).  
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AB018382; BAA76559.1; .  
DR EMBL; AL035637; CAB38509.1; .  
DR InterPro; IPR000461; Alpha\_amylase.  
DR Pfam; PF00128; alpha-amylase; 1  
KW Cell wall; transferase; Glycosyltransferase.  
FT CONFLICT 120 120 V -> VRRVMLLCSLTNKV (IN REF. 1).  
SQ SEQUENCE 2358 AA; 269192 MW; A175577C9D8AD731 CRC64;  
Query Match 38.8%; Score 45; DB 1; Length 2358;  
Best Local Similarity 50.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 7 ILEAWSNDPOYNK 20  
| | : | | | : | |  
69 ILDKWKDCDPRNE 82

RESULT 13  
HFBI\_HAEIN STANDARD; PRT; 241 AA.  
AC P35757;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE CHAPERONE PROTEIN HIFB PRECURSOR.  
GN HIFB.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ISOLATE R1369 / SEROTYPE B;  
RA Smith A.L., Forney L.J., Chanyangam M.;  
RA Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.  
RL [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AM30 (770235) / SEROTYPE B;  
RX MEDLINE=95089703; PubMed=7997179;  
RA van Ham M.S., van Alphen L., Mooi F.R., van Putten J.P.;  
RT "The fibrillar gene cluster of Haemophilus influenzae type b.";  
CC MOL. Microbiol. 13:673-684(1994).  
CC -1- FUNCTION: MEDIATES ASSEMBLY OF PILI BY FORMING SOLUBLE MULTIMERIC  
CC COMPLEXES WITH PILI SUBUNITS AS AN INTERMEDIATE STEP IN THE  
CC ASSEMBLY PROCESS. THIS PROTEIN IS INVOLVED IN TYPE B PILI (HIFA)  
CC ASSEMBLY.  
CC -1- SUBCELLULAR LOCATION: PERIPLASMIC.  
CC -1- SIMILARITY: BELONGS TO THE PERIPLASMIC PILUS CHAPERONE FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).

EMBL; X66606; CAA47175.1; -.  
 DR EMBL; X33502; CAA83901.1; -.  
 DR PIR; S24978.  
 DR InterPro; IPR001829; Pili\_chaperone.  
 DR Pfam; PF00345; pili\_assembly; 1.  
 DR PRINTS; PR00969; CHAPERONPILI.  
 DR ProDom; PD001447; Pili\_chaperone; 1.  
 DR PROSITE; PS00635; PILI\_CHAPERONE; 1.  
 KW Chaperone; Fimbria; Periplasmic; Signal; Immunoglobulin domain.  
 FT SIGNAL 1 27 POTENTIAL.  
 FT CHAIN 28 241 CHAPERONE PROTEIN HIFB.  
 FT VARIANT 17 21 MISSING (IN STRAIN AM30).  
 FT VARIANT 104 104 G -> S (IN STRAIN AM30).  
 SO SEQUENCE 241 AA; 26743 MW; AB7513BF9FC96D53 CRC64;

Query Match 38.4%; Score 44.5; DB 1; Length 241;  
 Best Local Similarity 40.0%; Pred. No. 9.9; Mismatches 5; Gaps 1;  
 Matches 8; Conservative 5; Indels 2; Indels 5; Gaps 1;

QY 6 SILEAWSNDN-----DPOYNK 20  
 ::::| | | | | | | | | |  
 Db 58 ALVQAWIDNGNPNADPKYTK 77

## RESULT 14

HFB2\_HAEIN  
 ID HFB2\_HAEIN STANDARD; PRT; 241 AA.  
 AC P45991;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE CHAPERONE PROTEIN HIFB PRECURSOR.  
 GN HIFB.

OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]

SEQUENCE FROM N.A.  
 RC STRAIN=86-0295 / LKP SPOTYPE 1;  
 RA Green B.A., Olmsted S.B.;  
 Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
 -1- FUNCTION: MEDIATES ASSEMBLY OF PILI BY FORMING SOLUBLE MULTIMERIC COMPLEXES WITH PILI SUBUNITS AS AN INTERMEDIATE STEP IN THE ASSEMBLY PROCESS. THIS PROTEIN IS INVOLVED IN TYPE B PILI (HIFA) ASSEMBLY.

-1- SUBCELLULAR LOCATION: PERIPLASMIC.  
 CC -1- SIMILARITY: BELONGS TO THE PERIPLASMIC PILUS CHAPERONE FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE DOMAIN.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).

-----  
 CC EMBL; U19730; AAA61815.1; ALT\_INIT.  
 DR InterPro; IPR001829; Pili\_chaperone.  
 DR Pfam; PF00345; pili\_assembly; 1.  
 DR PRINTS; PR00969; CHAPERONPILI.  
 DR ProDom; PD001447; Pili\_chaperone; 1.  
 DR PROSITE; PS00635; PILI\_CHAPERONE; 1.  
 KW Chaperone; Fimbria; Periplasmic; Signal; Immunoglobulin domain.  
 FT SIGNAL 1 27 POTENTIAL.  
 FT CHAIN 28 241 CHAPERONE PROTEIN HIFB.  
 SO SEQUENCE 241 AA; 26718 MW; 8BB6854A42A220F8 CRC64;

Query Match 38.4%; Score 44.5; DB 1; Length 241;  
 Best Local Similarity 40.0%; Pred. No. 9.9; Mismatches 5; Gaps 1;  
 Matches 8; Conservative 5; Indels 2; Indels 5; Gaps 1;

QY 6 SILEAWSNDN-----DPOYNK 20  
 ::::| | | | | | | | | |  
 Db 58 ALVQAWIDNGNPNADPKYTK 77

## RESULT 15

ROD1\_YEAST  
 ID ROD1\_YEAST STANDARD; PRT; 837 AA.  
 AC Q02805; Q12475;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE ROD1 PROTEIN.  
 GN ROD1 OR YOR018W OR OR26.08.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]

SEQUENCE FROM N.A.  
 RP MEDLINE=96216358; PubMed=8621680;  
 RX Wu A.L., Hallstrom T.C., Moye-Rowley W.S.;  
 "ROD1, a novel gene conferring multiple resistance phenotypes in  
 RT Saccharomyces cerevisiae."  
 RL J. Biol. Chem. 271:2914-2920(1996).  
 RN [2]

SEQUENCE FROM N.A.  
 RC STRAIN=S288C / FY1679;  
 RA de Haan M., Maarse A.C., Grivell L.A.;  
 Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
 RL -1- FUNCTION: MEDIATES RESISTANCE TO O-DINITROBENZENE, CALCIUM AND ZINC.

-1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED OR ASSOCIATED TO A COMPLEX THAT COFRAGMENTATION WITH PLASMA MEMBRANE.  
 CC -1- SIMILARITY: TO YEAST YFR022W, YGR068C AND TO S.POMBE SPAC31A2.12 AND SPAC8A4.13C. PARTIAL, TO D.DISCOIDEUM FILOPODIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).

-----  
 CC EMBL; U40561; AAB03678.1; -.  
 DR EMBL; X87331; CAA60767.1; -.  
 DR EMBL; Z74926; CAA99208.1; -.  
 DR SGD; S0005544; ROD1.  
 FT DOMAIN 3 6 POLY-SER.  
 FT CONFLICT 618 618 Y -> D (IN REF. 1).  
 SO SEQUENCE 837 AA; 92349 MW; FD9A64174BDC443 CRC64;

Query Match 37.9%; Score 44; DB 1; Length 837;  
 Best Local Similarity 38.9%; Pred. No. 50;  
 Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 3 NHI5LEAWSNDNDPOYNK 20  
 ::::| | | | | | | | | |  
 Db 574 SNLTISEWNNNSPSANR 591

Search completed: March 27, 2002, 14:27:03  
 Job time: 1645 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:26:04 ; Search time 188.53 Seconds  
(without alignments)  
16.293 Million cell updates/sec

Title: US-09-290-049A-11  
Perfect score: 116  
Sequence: 1 AINHL5ILEAWSNDNDPQYNKD 21  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_17.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	90.5	1575	2 Q9LCH3	Q9LCH3 streptococc
2	103	88.8	1577	2 Q54178	Q54178 streptococc
3	90	77.6	1577	2 Q5265	Q5265 streptococc
4	85	73.3	1449	2 Q68542	Q68542 streptococc
5	85	73.3	1449	2 Q55264	Q55264 streptococc
6	81	69.8	1338	2 Q9WXJ4	Q9WXJ4 streptococc
7	78	67.2	1477	2 Q9L466	Q9L466 leuconostoc
8	78	67.2	1508	2 Q52224	Q52224 leuconostoc
9	78	67.2	1508	2 Q9EZH5	Q9EZH5 leuconostoc
10	76	65.5	1512	2 Q9WXJ5	Q9WXJ5 streptococc
11	75	64.7	1390	2 Q69385	Q69385 streptococc
12	75	64.7	1455	2 Q69391	Q69391 streptococc
13	74	63.8	1590	2 Q55263	Q55263 streptococc
14	74	63.8	1590	2 Q59983	Q59983 streptococc
15	73	62.9	1016	2 Q9LCJ7	Q9LCJ7 leuconostoc
16	68	58.6	1518	2 Q00600	Q00600 streptococc
17	68	58.6	1527	2 Q9ZAR4	Q9ZAR4 leuconostoc
18	67	57.8	1599	2 Q00599	Q00599 streptococc
19	66	56.9	1455	2 Q69382	Q69382 streptococc

20	66	56.9	1455	2	Q69388	Q69388 streptococc
21	66	56.9	1455	2	Q69397	Q69397 streptococc
22	64	55.2	2057	2	Q9RE05	Q9RE05 leuconostoc
23	50	43.1	474	10	Q9SGV0	Q9SGV0 arabidopsis
24	49	42.2	286	2	Q9RDG2	Q9RDG2 streptomyc
25	49	42.2	632	4	Q9NTE0	Q9NTE0 homo sapien
26	49	42.2	759	4	Q9UGP8	Q9UGP8 homo sapien
27	49	42.2	760	4	Q95380	Q95380 homo sapien
28	47	40.5	175	10	Q9LMJ8	Q9LMJ8 arabidopsis
29	47	40.5	403	2	P96402	P96402 mycobacteri
30	47	40.5	418	12	Q85674	Q85674 reovirus sp
31	47	40.5	418	12	Q85675	Q85675 reovirus sp
32	47	40.5	418	12	Q85676	Q85676 reovirus sp
33	47	40.5	418	12	Q85677	Q85677 reovirus sp
34	47	40.5	418	12	Q85678	Q85678 reovirus sp
35	47	40.5	418	12	Q85681	Q85681 reovirus sp
36	47	40.5	418	12	Q85682	Q85682 reovirus sp
37	47	40.5	445	2	Q9RTN4	Q9RTN4 deinococcus
38	47	40.5	490	2	Q07382	Q07382 lactococcus
39	47	40.5	491	2	Q9CF88	Q9CF88 lactococcus
40	46	39.7	420	10	Q9T094	Q9T094 arabidopsis
41	46	39.7	460	10	Q9C5R0	Q9C5R0 arabidopsis
42	46	39.7	623	2	Q9XAJ3	Q9XAJ3 streptomyc
43	46	39.7	626	2	O51574	O51574 borrelia bu
44	46	39.7	688	12	Q83394	Q83394 mouse mamma
45	46	39.7	703	11	Q64698	Q64698 rattus norv

ALIGNMENTS

RESULT 1  
Q9LCH3 PRELIMINARY; PRT; 1575 AA.  
AC Q9LCH3; PRELIMINARY; PRT; 1575 AA.  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE.  
GN GTFR.  
OS Streptococcus oralis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1303;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC10557;  
RX MEDLINE=20231779; PubMed=10768934;  
RA Fujiwara T., Hoshino T., Ooshima T., Sobue S., Hamada S.;  
RT "Purification, characterization, and molecular analysis of the gene  
RT encoding glucosyltransferase from Streptococcus oralis.";  
RL Infect. Immun. 68:2475-2483(2000).  
DR EMBL; AB025228; BAA95201.1;  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding.1; 17.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
SQ TRANSFERASE  
SEQUENCE 1575 AA; 176792 MW; 772A26E4D7C2E543 CRC64;

Query Match 90.5%; Score 105; DB 2; Length 1575;  
Best Local Similarity 90.5%; Pred. No. 6.7e-08;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AINHL5ILEAWSNDNDPQYNKD 21  
|||  
Db 546 AIKHL5ILEAWSNDNDPQYNKD 566

RESULT 2  
Q54178 PRELIMINARY; PRT; 1577 AA.  
ID Q54178

Q54178; Q54247;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE.  
GN GTFG.  
OS Streptococcus gordonii challis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=29390;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CHALLIS;  
RX MEDLINE=92276337; PubMed=1534326;  
RA Sulavik M.C., Tardif G., Clewell D.B.;  
RT "Identification of a gene, rgg, which regulates expression of  
glucosyltransferase and influences the Spp phenotype of Streptococcus  
gordonii Challis.";  
RL J. Bacteriol. 174:3577-3586(1992).  
DR EMBL; U12643; AAC43483.1; -;  
DR EMBL; M89776; AAA26969.1; -;  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 18.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1577 AA; 177805 MW; 5AE0328DC5E08D18 CRC64;  
  
Query Match 88.8%; Score 103; DB 2; Length 1577;  
Best Local Similarity 85.7%; Pred. No. 1.4e-07;  
Matches 18; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 AINHLSILEAWSNDPQYNKD 21  
I: ||||| ||||| |||||  
Db 548 ALKHLISILEAWSNDPQYNKD 568  
  
RESULT 3  
AC Q55265; PRELIMINARY; PRT; 1577 AA.  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE PRECURSOR.  
GN GTFM.  
OS Streptococcus salivarius.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1304;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95122197; PubMed=7822030;  
RA Simpson C.L., Giffard P.M., Jacques N.A.;  
RT "Streptococcus salivarius ATCC 25975 possesses at least two genes  
coding for primer-independent glucosyltransferases.";  
RL Infect. Immun. 63:609-621(1995).  
DR EMBL; L35928; AAC41413.1; -;  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 11.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Signal; Transferase.  
FT SIGNAL 1 38 POTENTIAL.

FT CHAIN 39 1577 GLUCOSYLTRANSFERASE.  
SQ SEQUENCE 1577 AA; 175290 MW; 3EFB898A7D3A7BF3 CRC64;  
  
Query Match 77.6%; Score 90; DB 2; Length 1577;  
Best Local Similarity 85.7%; Pred. No. 1.6e-05;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 1 AINHLSILEAWSNDPQYNKD 21  
I: ||||| ||||| |||||  
Db 588 AIAHLSILEAWSNDPQYNKD 608  
  
RESULT 4  
AC Q68542; PRELIMINARY; PRT; 1449 AA.  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE N (FRAGMENT).  
GN GTFN.  
OS Streptococcus salivarius.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1304;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VI477;  
RA Jaffe R.I.;  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBDJ databases.  
DR EMBL; AF049609; AAC05156.1; -;  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 8.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
FT NON\_TER 1449 1449  
SQ SEQUENCE 1449 AA; 159895 MW; 0700F6D748471BFB CRC64;  
  
Query Match 73.3%; Score 85; DB 2; Length 1449;  
Best Local Similarity 76.2%; Pred. No. 8.8e-05;  
Matches 16; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
  
Qy 1 AINHLSILEAWSNDPQYNKD 21  
I: ||||| ||||| |||||  
Db 536 AIKHLISILEAWSNDAYNED 556  
  
RESULT 5  
AC Q55264; PRELIMINARY; PRT; 1449 AA.  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE PRECURSOR.  
GN GTFL.  
OS Streptococcus salivarius.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1304;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95122197; PubMed=7822030;  
RA Simpson C.L., Giffard P.M., Jacques N.A.;  
RT "Streptococcus salivarius ATCC 25975 possesses at least two genes  
coding for primer-independent glucosyltransferases.";  
RL Infect. Immun. 63:609-621(1995).  
DR EMBL; L35495; AAC41412.1; -;  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.

DR Pfam; PF01473; CW\_binding\_1; 8.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Signal; Transferase.  
FT SIGNAL 1 35 POTENTIAL.  
FT CHAIN 36 1449 GLUCOSYLTRANSFERASE.  
SQ SEQUENCE 1449 AA; 159984 MW; DD62F07306E86A46 CRC64;

Query Match 73.3%; Score 85; DB 2; Length 1449;  
Best Local Similarity 76.2%; Pred. No. 8.8e-05;  
Matches 16; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 AINHLSILEAWSNDPQYNKD 21  
ID 1 AINHLSILEAWSNDPQYNKD 21  
Db 536 AIKHLSTLEAWSNDAYNED 556

RESULT 6  
Q9WXJ4 PRELIMINARY; PRT; 1338 AA.  
ID Q9WXJ4  
Q9WXJ4;  
01-NOV-1999 (TrEMBLrel. 12, Created)  
01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
GTF-S.  
GN GTF.  
OS Streptococcus criceti.  
OG Plasmid pAM1.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1333;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HS-6;  
RA Inoue M., Fukui K., Miyagi A.;  
RT "S.cricetus glucosyltransferase(gtfs and gtf) genes."  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB026123; BAA77236.1; -.  
DR HSSP; P06278; 1VJS.  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 10.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Plasmid.  
SQ SEQUENCE 1338 AA; 148558 MW; 0A90C8E10E15D99B CRC64;

Query Match 69.8%; Score 81; DB 2; Length 1338;  
Best Local Similarity 76.2%; Pred. No. 0.00034;  
Matches 16; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 AINHLSILEAWSNDPQYNKD 21  
ID 1 AINHLSILEAWSNDPQYNKD 21  
Db 437 AIDHLSILEAWSNDYVKD 457

RESULT 7  
Q9L466 PRELIMINARY; PRT; 1477 AA.  
ID Q9L466  
Q9L466;  
01-OCT-2000 (TrEMBLrel. 15, Created)  
01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DEXTRANSUCRASE (EC 2.4.1.5).  
DSRC.  
GN DSR.  
OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OC Leuconostoc.  
OX NCBI\_TaxID=1245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NRRL B-1355;  
RA Arguello-Morales M.A., Renaud-Simeon M., Pizzut S., Sarcabal P.,

RA Willemot R.M., Monsan P.;  
RT "Sequence analysis of the gene encoding alternansucrase, a sucrose  
RL glucosyltransferase from Leuconostoc mesenteroides NRRL B-1355.";  
Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ250172; CAB76565.1; -.  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 14.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase; Glycosyltransferase.  
SQ SEQUENCE 1477 AA; 164886 MW; E6F5710DEDFCB831 CRC64;

Query Match 67.2%; Score 78; DB 2; Length 1477;  
Best Local Similarity 71.4%; Pred. No. 0.0011;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 AINHLSILEAWSNDPQYNKD 21  
ID 1 AINHLSILEAWSNDPQYNKD 21  
Db 532 ANQHLSILEDWSHNDPEYVKD 552

RESULT 8  
O52224 PRELIMINARY; PRT; 1508 AA.  
ID O52224  
O52224;  
01-JUN-1998 (TrEMBLrel. 06, Created)  
01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DEXTRANSUCRASE (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE 6-  
GLUCOSYLTRANSFERASE).  
GN DSRB.  
OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OC Leuconostoc.  
OX NCBI\_TaxID=1245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NRRL B-1299;  
RA Monchois V., Renaud-Simeon M., Monsan P., Willemot R.M.;  
RL FEMS Microbiol. Lett. 0:0-0(1998).  
CC -1- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) = D-  
FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).  
DR EMBL; AF030129; AAB95453.1; -.  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 14.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase; Glycosyltransferase.  
SQ SEQUENCE 1508 AA; 168511 MW; E70CECB57A70D1F0 CRC64;

Query Match 67.2%; Score 78; DB 2; Length 1508;  
Best Local Similarity 71.4%; Pred. No. 0.0012;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 AINHLSILEAWSNDPQYNKD 21  
ID 1 AINHLSILEAWSNDPQYNKD 21  
Db 563 ANQHLSILEDWSHNDPEYVKD 583

RESULT 9  
Q9EZH5 PRELIMINARY; PRT; 1508 AA.  
ID Q9EZH5  
Q9EZH5;  
01-MAR-2001 (TrEMBLrel. 16, Created)  
01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DEXTRANSUCRASE DSRB742.  
GN DSRB742.  
OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OC Leuconostoc.





DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE GTF-I.  
OS Streptococcus sobrinus.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1310;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 33478;  
RA Sato S.;  
RL Ann. Kagoshima Univ. Dental School 16:23-29(1996).  
DR EMBL; D63570; BAA09792.1; -.  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 15.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1590 AA; 176057 MW; 9DF7A3F2C6E4FD43 CRC64;

Query Match 63.8%; Score 74; DB 2; Length 1590;  
Best Local Similarity 66.7%; Pred. No. 0.0053;  
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 AINHLSTLEAWSNDNDPOYNKD 21  
| | | | | | | | | | | | | | | | | | | | | |  
Db 477 ANHVSIVEAWSNDNDTPYLHD 497

RESULT 14  
ID Q59983 PRELIMINARY; PRT; 1590 AA.  
AC Q59983;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE  
DE 6-GLUCOSYLTRANSFERASE).  
GN GTFI.  
OS Streptococcus sobrinus.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1310;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OMZ176;  
RY MEDLINE=94146405; PubMed=8312602;  
Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;  
"DNA sequence of the glucosyltransferase gene of serotype d  
Streptococcus sobrinus.";  
RL DNA Seq. 4:19-27(1993).  
CC -1- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) = D-  
CC FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).  
DR EMBL; D13858; BAA02976.1; -.  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 16.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Signal; Transferase; Glycosyltransferase.  
FT SIGNAL 1 38 POTENTIAL.  
FT CHAIN 39 1590 GLUCOSYLTRANSFERASE-I.  
SQ SEQUENCE 1590 AA; 175955 MW; C3C83A57CF3C2B0E CRC64;

Query Match 63.8%; Score 74; DB 2; Length 1590;  
Best Local Similarity 66.7%; Pred. No. 0.0053;  
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 AINHLSTLEAWSNDNDPOYNKD 21  
| | | | | | | | | | | | | | | | | | | | | |  
Db 477 ANHVSIVEAWSNDNDTPYLHD 497

RESULT 15  
Q9LCJ7 PRELIMINARY; PRT; 1016 AA.  
ID Q9LCJ7  
AC Q9LCJ7;  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE DEXTRANSUCRASE.  
GN DSRT.  
OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OC Leuconostoc.  
OX NCBI\_TaxID=1245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NRRL B-512F;  
RX MEDLINE=20169623; PubMed=10705445;  
RA Funane K., Mizuno K., Takahara H., Kobayashi M.;  
"Gene encoding a dextranucrase-like protein in Leuconostoc  
mesenteroides NRRL B-512F.";  
RL Biosci. Biotechnol. Biochem. 64:29-38(2000).  
DR EMBL; AB020020; BAA90527.1; -.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
SQ SEQUENCE 1016 AA; 110343 MW; 8896EFDE13CCCB47 CRC64;

Query Match 62.9%; Score 73; DB 2; Length 1016;  
Best Local Similarity 72.2%; Pred. No. 0.0047;  
Matches 13; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 HLSILEAWSNDNDPOYNKD 21  
| | | | | | | | | | | | | | | | | | | | | |  
Db 551 HVSILEDWSDNDABEYVKD 568

Search completed: March 27, 2002, 14:26:04  
Job time: 1675 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:57:55 ; Search time 198.55 Seconds  
(without alignments)  
7.834 Million cell updates/sec

Title: US-09-290-049a-12

Perfect score: 119

Sequence: 1 ANHVSIVEAWSNDTPYLHD 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

1 number of hits satisfying chosen parameters: 522463

Maximum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_1101:\*

- 1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT:\*
- 2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT:\*
- 3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT:\*
- 4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT:\*
- 5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT:\*
- 6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT:\*
- 7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT:\*
- 8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT:\*
- 9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT:\*
- 10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT:\*
- 11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT:\*
- 12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT:\*
- 13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT:\*
- 14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT:\*
- 15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT:\*
- 16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT:\*
- 17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT:\*
- 18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT:\*
- 19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT:\*
- 20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT:\*
- 21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT:\*
- 22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119	100.0	1592	14	AA1980.DAT:*
2	57	47.9	1577	17	AA1981.DAT:*
3	55	46.2	2057	21	AA1982.DAT:*
4	53	44.5	195	21	AA1983.DAT:*
5	51	42.9	1195	20	AA1984.DAT:*
6	48.5	40.8	401	22	AA1985.DAT:*
7	46	38.7	486	22	AA1986.DAT:*
8	45	37.8	174	21	AA1987.DAT:*
9	45	37.8	185	21	AA1988.DAT:*
10	45	37.8	189	21	AA1989.DAT:*
11	45	37.8	199	22	AA1990.DAT:*

12	45	37.8	305	21	AA1980.DAT:*
13	45	37.8	339	21	AA1981.DAT:*
14	45	37.8	603	18	AA1982.DAT:*
15	45	37.8	1095	18	AA1983.DAT:*
16	45	37.8	1366	21	AA1984.DAT:*
17	43.5	36.6	202	16	AA1985.DAT:*
18	43.5	36.6	1092	17	AA1986.DAT:*
19	43.5	36.6	1092	20	AA1987.DAT:*
20	43	36.1	88	21	AA1988.DAT:*
21	43	36.1	151	20	AA1989.DAT:*
22	43	36.1	160	22	AA1990.DAT:*
23	43	36.1	163	22	AA1991.DAT:*
24	43	36.1	220	22	AA1992.DAT:*
25	43	36.1	253	22	AA1993.DAT:*
26	43	36.1	256	22	AA1994.DAT:*
27	43	36.1	271	21	AA1995.DAT:*
28	43	36.1	271	22	AA1996.DAT:*
29	43	36.1	304	20	AA1997.DAT:*
30	43	36.1	437	21	AA1998.DAT:*
31	43	36.1	541	20	AA1999.DAT:*
32	42	35.3	146	18	AA2000.DAT:*
33	42	35.3	201	21	AA2001.DAT:*
34	42	35.3	334	11	AA2002.DAT:*
35	42	35.3	455	22	AA2003.DAT:*
36	42	35.3	530	18	AA2004.DAT:*
37	42	35.3	540	21	AA2005.DAT:*
38	42	35.3	540	21	AA2006.DAT:*
39	42	35.3	775	21	AA2007.DAT:*
40	42	35.3	775	21	AA2008.DAT:*
41	42	35.3	1627	12	AA2009.DAT:*
42	42	35.3	1627	15	AA2010.DAT:*
43	42	35.3	1627	16	AA2011.DAT:*
44	42	35.3	3898	12	AA2012.DAT:*
45	41.5	34.9	325	22	AA2013.DAT:*

#### ALIGNMENTS

RESULT 1	
AA1980.DAT:*	
ID AAR32925 standard; Protein; 1592 AA.	
XX	
AC AAR32925;	
XX	
DT 28-JUN-1993 (first entry)	
XX	
DE Glucosyltransferase I.	
XX	
KW GT-1; Streptococcus; dental; caries.	
XX	
OS Streptococcus sobrinus.	
XX	
PN JP05023180	
XX	
PD 02-FEB-1993	
XX	
PF 25-JUL-1993; 91JP-0186592.	
XX	
PR 25-JUL-1993; 91JP-0186592.	
XX	
PA (FUKU)/ FUKUI I.	
PA (KATO)/ KATO K.	
XX	
XX	
DR WPI; 1993-079449/10.	
DR N-PSDB; AAQ37760.	
XX	
PT DNA sequence glucosyltransferase-I - comprises Streptococcus	
PT sobrinus DNA sequence with at least one nucleotide added or	
XX deleted	
PS Claim 13; Page 15; 29pp; Japanese.	
XX	

Arabidopsis thalia  
Arabidopsis thalia  
Photobacterium lum  
Photobacterium lum  
Human ORF1553  
Morcellia sp. alph  
Algal alpha-1.4-gl  
Gracilariaopsis lem  
Human secreted pro  
Chlamydia pneumoni  
Human membrane-bou  
Human membrane-bou  
S. epidermidis ope  
Human membrane-bou  
Human membrane-bou  
Amino acid sequenc  
Human membrane-bou  
Human prostate can  
Human secreted pro  
Chlamydia trachoma  
Staphylococcus aur  
Human ADAM protein  
Amino acid sequenc  
Amino acid sequenc  
Human metalloprote  
Human ADAM protein  
Human ADAM protein  
Human ADAM protein  
Human ADAM protein  
M. pneumoniae Pl c  
Mycoplasma pneumon  
Cytadhesin Pl. My  
Hog cholera virus  
Human polypeptide





PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 26-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 44.5%; Score 53; DB 21; Length 195;  
 Best Local Similarity 42.1%; Pred. No. 2.2;  
 Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 3 NHVSIVEAWSNDTPYLHD 21  
 || :|||:|  
 Db 109 nhqvidawdsdqkplwrd 127  
 RESULT 5

AAW75420  
 ID AAW75420 standard; protein; 1195 AA.  
 XX  
 AC AAW75420;  
 XX  
 DT 16-MAR-1999 (first entry)  
 XX  
 DE T.thermophilus nitrate reductase alpha subunit.  
 XX  
 KW Heat-stable; nitrate reductase; temperature; detection; food; toxicity;  
 KW carcinogen.  
 XX  
 OS Thermus thermophilus.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 630 /label= unknown  
 FT Misc-difference 669 /label= unknown  
 FT Misc-difference 691 /label= unknown  
 FT  
 XX ES2121561-A1.  
 XX  
 XX 16-NOV-1998.  
 PD  
 XX 09-MAY-1997; 97ES-0001003.  
 PF  
 XX 09-MAY-1997; 97ES-0001003.  
 PR  
 XX (UYMA-) UNIV AUTONOMA MADRID.  
 PA  
 XX WPI; 1999-001909/01.  
 DR  
 XX Heat stable nitrate reductase for high temperature nitrate detection  
 PT - comprises Thermus thermophilus derivative enhancing nitrite or  
 PT nitrate reduction  
 XX  
 PS Disclosure; Fig 2; 8pp; Spanish.  
 XX  
 CC This sequence represents the amino acid sequence of the Thermus  
 CC thermophilus heat-stable nitrate reductase alpha subunit. Heat stable  
 CC nitrate reductase can be used for high-temperature detection of nitrates  
 CC in samples, e.g. in food, where high levels of nitrates can be toxic or  
 CC carcinogenic.  
 XX  
 SQ Sequence 1195 AA;  
 Query Match 42.9%; Score 51; DB 20; Length 1195;  
 Best Local Similarity 42.9%; Pred. No. 37;  
 Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;  
 Qy 1 ANNHVSIVEAWSNDTPYLHD 21  
 ||| :||:|  
 Db 293 avnhllkefyadvrepyfqd 313  
 RESULT 6  
 AAM43544  
 ID AAM43544 standard; Protein; 401 AA.  
 XX  
 AC AAM43544;  
 XX  
 DT 22-OCT-2001 (first entry)  
 XX  
 DE Human polypeptide SEQ ID NO 222.  
 XX  
 KW Human; antiarthritic; antirheumatic; antiproliferative; vasotropic;  
 KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;  
 KW fungicide; ophthalmologic; cytostatic; immunosuppressive; nootropic;  
 KW neuroprotective; antiallergic; hepatotropic; antidiabetic;  
 KW antiinflammatory; antiulcer; vulnerary; anticonvulsant; antibacterial;

KW antiparasitic; cardiant; gene therapy; cancer; immune disorder;  
KW cardiovascular disorder; neurological disease; infection; human.  
XX Homo sapiens.  
XX WO200155308-A2.  
XX 02-AUG-2001.  
PD 17-JAN-2001; 2001WO-US01309.  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0241826.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 05-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251889.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.

PR 05-JAN-2001; 2001US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
PI WPI; 2001-488781/53.  
DR N-PSDB; AAI63850.  
XX New isolated nucleic acids and polypeptides, useful for diagnosing,  
PT treating and/or preventing human diseases and disorders -  
XX  
XX Claim 11; SEQ ID NO 222; 664pp + Sequence Listing; English.  
XX  
XX The invention relates to human polynucleotides (AAI63803-AAI64012) and  
CC the encoded proteins (AAI63803-AAI64012) useful for preventing, treating  
CC or ameliorating medical conditions e.g. by protein or gene therapy. The  
CC genes were isolated from a range of human tissues disclosed in the  
CC specification. The nucleic acids, proteins, antibodies and (ant)agonists  
CC are useful in the diagnosis, treatment and prevention of: (a) cancer,  
CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,  
CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or  
CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,  
CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,  
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 401 AA;

Query Match 40.8%; Score 48.5; DB 22; Length 401;  
Best Local Similarity 47.4%; Pred. No. 25;  
Matches 9; Conservative 3; Mismatches 4; Indels 3; Gaps 1;  
QY 5 VSIVAWSDN---DTPYLH 20  
Db 46 vallevkdnrtdxpxlh 64  
|:::| | | | | | |  
|:::| | | | | | |  
RESULT 7  
36552  
AAB36552 standard; Protein; 486 AA.  
AC AAB36552;  
XX  
XX DT 07-MAR-2001 (first entry)  
XX  
XX DE Lawsonia intracellularis flagellar hook protein FlgE SEQ ID NO:1.  
XX  
XX DE Lawsonia intracellularis; flgE; flagellar hook protein; vaccine;  
KW intestinal disease; immunogenic; diagnosis; antibacterial; swine;  
KW pig; infection; detection; identification.  
XX  
XX OS Lawsonia intracellularis.  
XX  
XX PN WO200069904-A1.  
XX  
XX PD 23-NOV-2000.  
XX  
XX PF 11-MAY-2000; 2000WO-AU00437.  
XX  
XX PR 13-MAY-1999; 99US-0133973.  
XX  
XX (PFIZ ) PFIZER PROD INC.  
PA (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.  
PA (PIGR-) PIG RES & DEV CORP.  
XX

PI Panaccio M, Rosey EL, Sinistaj M, Hasse D, Parsons J;  
PI Ankenbauer RG;  
XX  
XX WPI; 2001-016210/02.  
DR N-PSDB; AAC88037.  
XX  
XX New immunogenic Lawsonia FlgE peptide, its nucleic acid and antibody,  
PT useful in vaccines and diagnosis of Lawsonia infections, particularly  
PT in swine -  
XX  
XX Claim 13; Page 87-90; 97pp; English.  
PS  
XX  
XX The present sequence is the Lawsonia intracellularis flagellar hook  
CC protein FlgE. The present invention describes an isolated or recombinant  
CC polypeptide (I) that comprises, mimics or cross-reacts with a B- or  
CC T-cell epitope of a FlgE (flagellar hook) polypeptide from a  
CC Lawsonia spp. (I) has antibacterial activity, and induces a specific  
CC humoral immune response. (I) are used as antigens in vaccines to prevent  
CC or treat infection by Lawsonia, in birds and animals, especially pigs,  
CC to raise specific antibodies (Ab) and to detect past or present  
CC infection. Ab are also useful in diagnosis, to detect L. intracellularis  
CC or immunologically cross-reactive species, also for identification of  
CC epitopes in FlgE. Vectors that contain nucleic acids (II) encoding (I)  
CC are also useful in genetic vaccines, and fragments of (II) are useful  
CC as primers or probes for detecting L. intracellularis or related  
CC microorganisms, in hybridisation or amplification assays.  
XX  
XX Sequence 486 AA;

Query Match 38.7%; Score 46; DB 22; Length 486;  
Best Local Similarity 41.2%; Pred. No. 73;  
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;  
QY 1 ANNHVSIVEAWSNDNDTP 17  
Db 169 anpyfalleeswkgngtp 185  
| | : : : | | | | |  
| | : : : | | | | |  
RESULT 8  
AAG60587  
ID AAG60587 standard; Protein; 174 AA.  
XX  
XX AC AAG60587;  
XX  
XX DT 18-OCT-2000 (first entry)  
XX  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 78492.  
XX  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
XX OS Arabidopsis thaliana.  
XX  
XX PN EP1033405-A2.  
XX  
XX PD 06-SEP-2000.  
XX  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX PR 25-FEB-1999; 99US-0121825.  
XX  
XX PR 05-MAR-1999; 99US-0123180.  
XX  
XX PR 09-MAR-1999; 99US-0123548.  
XX  
XX PR 23-MAR-1999; 99US-0125788.  
XX  
XX PR 25-MAR-1999; 99US-0126264.  
XX  
XX PR 29-MAR-1999; 99US-0126785.  
XX  
XX PR 01-APR-1999; 99US-0127462.  
XX  
XX PR 06-APR-1999; 99US-0128234.  
XX  
XX PR 08-APR-1999; 99US-0128714.  
XX  
XX PR 16-APR-1999; 99US-0129845.  
XX  
XX PR 19-APR-1999; 99US-0130077.  
XX  
XX PR 21-APR-1999; 99US-0130449.



PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 25-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144432.  
PR 20-JUL-1999; 99US-0144435.  
PR 20-JUL-1999; 99US-01444352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 28-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.

PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 21-OCT-1999; 99US-0160880.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 37.8%; Score 45; DB 21; Length 174;  
Best Local Similarity 50.0%; Pred. No. 31;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 5 VSIVEANSNDTPY 18

Db 122 vyvteswgsstpy 135

#### RESULT 9

AG60586  
ID AAG60586 standard; Protein: 185 AA.

XX AC AAG60586;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 78491.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX AC Arabidopsis thaliana.

XX EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.  
XX PR 05-MAR-1999; 99US-0123180.  
XX PR 09-MAR-1999; 99US-0123548.  
XX PR 23-MAR-1999; 99US-0125788.  
XX PR 25-MAR-1999; 99US-0126284.  
XX PR 29-MAR-1999; 99US-0126785.  
XX PR 01-APR-1999; 99US-0127462.  
XX PR 06-APR-1999; 99US-0128234.  
XX PR 08-APR-1999; 99US-0128714.  
XX PR 16-APR-1999; 99US-0129845.  
XX PR 19-APR-1999; 99US-0130077.  
XX PR 21-APR-1999; 99US-0130449.  
XX PR 23-APR-1999; 99US-0130510.  
XX PR 23-APR-1999; 99US-0130891.  
XX PR 28-APR-1999; 99US-0131449.  
XX PR 30-APR-1999; 99US-0132048.  
XX PR 30-APR-1999; 99US-0132407.  
XX PR 04-MAY-1999; 99US-0132484.  
XX PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.



PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136780.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 26-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 13-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.

Sequence 199 AA;

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 21-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145098.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 37.8%; Score 45; DB 21; Length 305;  
Best Local Similarity 47.1%; Pred. No. 60;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
QY 3 NHVSIVEAWSNDTPYL 19

```
Db      200 dyegimeawskgtlyv 216
      :: |:|||| | |:
RESULT 13
AAG09844
ID      AAG09844 standard; Protein; 339 AA.
XX
AC      AAG09844;
XX
DT      17-OCT-2000 (first entry)
XX
DE      Arabidopsis thaliana protein fragment SEQ ID NO: 7934.
XX
KW      Protein identification; signal transduction pathway; metabolic pathway;
KW      hybridisation assay; genetic mapping; gene expression control; promoter;
KW      termination sequence.
XX
OS      Arabidopsis thaliana.
XX
EP      EP1033405-A2.
XX
PE      06-SEP-2000.
XX
PF      25-FEB-2000; 2000EP-0301439.
XX
PR      25-FEB-1999; 99US-0121825.
PR      05-MAR-1999; 99US-0123180.
PR      09-MAR-1999; 99US-0123548.
PR      23-MAR-1999; 99US-0125788.
PR      25-MAR-1999; 99US-0126264.
PR      29-MAR-1999; 99US-0126785.
PR      01-APR-1999; 99US-0127462.
PR      06-APR-1999; 99US-0128234.
PR      08-APR-1999; 99US-0128714.
PR      16-APR-1999; 99US-0129845.
PR      21-APR-1999; 99US-0130077.
PR      23-APR-1999; 99US-0130449.
PR      28-APR-1999; 99US-0130891.
PR      30-APR-1999; 99US-0131449.
PR      30-APR-1999; 99US-0132048.
PR      04-MAY-1999; 99US-0132407.
PR      05-MAY-1999; 99US-0132484.
PR      06-MAY-1999; 99US-0132486.
PR      06-MAY-1999; 99US-0132487.
PR      07-MAY-1999; 99US-0132863.
PR      11-MAY-1999; 99US-0134256.
PR      14-MAY-1999; 99US-0134218.
PR      14-MAY-1999; 99US-0134219.
PR      14-MAY-1999; 99US-0134221.
PR      18-MAY-1999; 99US-0134370.
PR      19-MAY-1999; 99US-0134768.
PR      20-MAY-1999; 99US-0134941.
PR      21-MAY-1999; 99US-0135124.
PR      24-MAY-1999; 99US-0135353.
PR      25-MAY-1999; 99US-0135629.
PR      27-MAY-1999; 99US-0136021.
PR      28-MAY-1999; 99US-0136392.
PR      01-JUN-1999; 99US-0136782.
PR      03-JUN-1999; 99US-0137222.
PR      04-JUN-1999; 99US-0137528.
PR      07-JUN-1999; 99US-0137502.
PR      08-JUN-1999; 99US-0137724.
PR      10-JUN-1999; 99US-0138094.
PR      10-JUN-1999; 99US-0138540.
PR      14-JUN-1999; 99US-0138847.
PR      16-JUN-1999; 99US-0139119.
PR      16-JUN-1999; 99US-0139452.
PR      17-JUN-1999; 99US-0139453.
PR      18-JUN-1999; 99US-0139492.
PR      18-JUN-1999; 99US-0139454.
PR      18-JUN-1999; 99US-0139455.
PR      18-JUN-1999; 99US-0139456.
PR      18-JUN-1999; 99US-0139457.
PR      18-JUN-1999; 99US-0139458.
PR      18-JUN-1999; 99US-0139459.
PR      18-JUN-1999; 99US-0139460.
PR      18-JUN-1999; 99US-0139461.
PR      18-JUN-1999; 99US-0139462.
PR      18-JUN-1999; 99US-0139463.
PR      18-JUN-1999; 99US-0139750.
PR      18-JUN-1999; 99US-0139763.
PR      21-JUN-1999; 99US-0139817.
PR      22-JUN-1999; 99US-0139899.
PR      23-JUN-1999; 99US-0140353.
PR      23-JUN-1999; 99US-0140354.
PR      24-JUN-1999; 99US-0140695.
PR      28-JUN-1999; 99US-0140823.
PR      29-JUN-1999; 99US-0140991.
PR      30-JUN-1999; 99US-0141287.
PR      01-JUL-1999; 99US-0141842.
PR      01-JUL-1999; 99US-0142154.
PR      02-JUL-1999; 99US-0142055.
PR      06-JUL-1999; 99US-0142390.
PR      08-JUL-1999; 99US-0142803.
PR      09-JUL-1999; 99US-0142920.
PR      12-JUL-1999; 99US-0142977.
PR      13-JUL-1999; 99US-0143542.
PR      14-JUL-1999; 99US-0143624.
PR      15-JUL-1999; 99US-0144005.
PR      16-JUL-1999; 99US-0144085.
PR      16-JUL-1999; 99US-0144086.
PR      19-JUL-1999; 99US-0144325.
PR      19-JUL-1999; 99US-0144331.
PR      19-JUL-1999; 99US-0144332.
PR      19-JUL-1999; 99US-0144333.
PR      19-JUL-1999; 99US-0144334.
PR      20-JUL-1999; 99US-0144352.
PR      20-JUL-1999; 99US-0144632.
PR      20-JUL-1999; 99US-0144884.
PR      21-JUL-1999; 99US-0144814.
PR      21-JUL-1999; 99US-0145086.
PR      21-JUL-1999; 99US-0145088.
PR      22-JUL-1999; 99US-0145085.
PR      22-JUL-1999; 99US-0145087.
PR      22-JUL-1999; 99US-0145089.
PR      22-JUL-1999; 99US-0145192.
PR      23-JUL-1999; 99US-0145145.
PR      23-JUL-1999; 99US-0145218.
PR      23-JUL-1999; 99US-0145224.
PR      26-JUL-1999; 99US-0145276.
PR      27-JUL-1999; 99US-0145913.
PR      27-JUL-1999; 99US-0145918.
PR      27-JUL-1999; 99US-0145919.
PR      28-JUL-1999; 99US-0145951.
PR      02-AUG-1999; 99US-0146386.
PR      02-AUG-1999; 99US-0146388.
PR      02-AUG-1999; 99US-0146389.
PR      04-AUG-1999; 99US-0147038.
PR      04-AUG-1999; 99US-0147204.
PR      05-AUG-1999; 99US-0147302.
PR      05-AUG-1999; 99US-0147192.
PR      05-AUG-1999; 99US-0147260.
PR      06-AUG-1999; 99US-0147303.
PR      06-AUG-1999; 99US-0147416.
PR      09-AUG-1999; 99US-0147493.
PR      09-AUG-1999; 99US-0147935.
PR      10-AUG-1999; 99US-0148171.
PR      11-AUG-1999; 99US-0148319.
PR      12-AUG-1999; 99US-0148341.
PR      13-AUG-1999; 99US-0148565.
PR      13-AUG-1999; 99US-0148684.
PR      16-AUG-1999; 99US-0149368.
```

```

PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149829.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 37.8%; Score 45; DB 21; Length 339;
Best Local Similarity 47.1%; Pred. No. 68;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 3 NHVSIVEAWSNDTPYL 19
   :|:||||| | |
Db 234 dyeqimeawsdkgtlyv 250

RESULT 14
AAWI7889
ID AAWI7889 standard; Protein; 603 AA.
XX

PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149829.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 37.8%; Score 45; DB 21; Length 339;
Best Local Similarity 47.1%; Pred. No. 68;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 3 NHVSIVEAWSNDTPYL 19
   :|:||||| | |
Db 234 dyeqimeawsdkgtlyv 250

RESULT 14
AAWI7889
ID AAWI7889 standard; Protein; 603 AA.
XX

XX AAWI7889;
XX AC.
XX DT 29-JAN-1998 (first entry)
XX DE Photorhabdus luminescens insect toxin protein TcaA11.
XX XX
XX KW Insecticide; insect; toxin; pest control; biological control;
XX KW Photorhabdus luminescens; TcaA; Southern corn rootworm;
XX KW Colorado potato beetle; Western corn rootworm; meal worm;
XX KW boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;
XX KW cabbage looper; codling moth; corn earworm; European corn borer;
XX KW tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;
XX KW Diptera, Dictyoptera; Acarina; Homoptera.
XX OS Photorhabdus luminescens strain W-14 (ATCC 55397).
XX PN WO9717432-A1.
XX XX
XX PD 15-MAY-1997.
XX PF 06-NOV-1996; 96WO-US18003.
XX PR 28-AUG-1996; 96US-0705484.
XX PR 06-NOV-1995; 95US-0007255.
XX PR 28-FEB-1996; 96US-0608423.
XX PA (WISC ) WISCONSIN ALUMNI RES FOUND.
XX PI Blackburn MB, Bowen DJ, Ciche TA, Ensign JC, Fatig R;
XX PI Ffrench-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;
XX PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;
XX PI Strickland JA;
XX XX
XX DR WPI; 1997-281022/25.
XX DR N-PSDB; AAT68841.
XX XX
XX PT Photorhabdus sp. insecticidal protein toxins and DNA encoding them -
XX PT can be genetically engineered into insect larvae food and plants for
XX PT insect control
XX PS Claim 34; Page 177-179; 276pp; English.
XX XX
XX CC This polypeptide comprises a claimed insecticidal toxin protein,
XX CC TcaA11 (see also AAWI7888), of Photorhabdus luminescens that
XX CC is a component of a toxin protein complex. Claimed toxins of P.
XX CC luminescens (see AAWI7871, AAWI7884-89, AAWI7899-900, AAWI8301-06) can
XX CC be applied to, or genetically engineered into, insect larvae food and
XX CC plants for insect control. The Photorhabdus toxins are particularly
XX CC effective against Southern corn rootworm, Colorado potato beetle,
XX CC Western corn rootworm, meal worm, boll weevil and turf grub
XX CC (Coleoptera), beet armyworm, black cutworm, cabbage looper, codling
XX CC moth, corn earworm, European corn borer, tobacco hornworm and
XX CC tobacco budworm (Lepidoptera), and are also active against insects
XX CC of the orders Hymenoptera, Diptera, Dictyoptera, Acarina and
XX CC Homoptera. (All claimed).
XX SQ Sequence 603 AA;

Query Match 37.8%; Score 45; DB 18; Length 603;
Best Local Similarity 33.3%; Pred. No. 1.3e+02;
Matches 7; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 ANNHVSIVEAWSNDTPYLHD 21
   :|: | | | | | |
Db 395 sdnpftlanywhdnqtlfshd 415

RESULT 15
AAWI7888
ID AAWI7888 standard; Protein; 1095 AA.
XX

```



Db 887 sdnprftlanyw hdnatlfshd 907

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:59:30 ; Search time 87.3 Seconds  
(without alignments)  
5.413 Million cell updates/sec

Title: US-09-290-049A-12  
Perfect score: 119  
Sequence: 1 ANNHVSIVEAWSNDTPYLHD 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

al number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	108	90.8	1475	3	US-09-007-999-2
2	108	90.8	1475	4	US-09-210-361-2
3	99	83.2	1375	4	US-09-210-361-4
4	74	62.2	1430	3	US-09-008-172-2
5	74	62.2	1430	4	US-09-210-361-6
6	57	47.9	1577	2	US-08-793-824-2
7	42	35.3	200	2	US-08-836-442-4
8	42	35.3	334	6	5290690-11
9	42	35.3	455	4	US-09-362-473-6
10	42	35.3	529	2	US-08-836-442-3
11	42	35.3	1627	1	US-07-665-792E-9
12	42	35.3	3898	4	US-08-750-717-2
13	41	34.5	259	1	US-08-277-231A-3
14	41	34.5	259	2	US-08-473-750-6
15	41	34.5	259	2	US-08-477-326-6
16	41	34.5	2465	2	US-08-596-291-3
17	41	34.5	2465	3	US-09-100-804-3
18	41	34.5	2466	3	US-09-080-855-12
19	41	34.5	2466	5	PCT-US94-09943-2
20	41	34.5	2485	4	US-09-290-640-46
21	41	34.5	3898	2	US-08-876-991-2
22	41	34.5	3898	2	US-09-059-853-2
23	40.5	34.0	214	4	US-09-214-278-1
24	40.5	34.0	1055	4	US-09-214-278-2
25	40.5	34.0	1148	4	US-08-882-046-4
26	40.5	34.0	1212	4	US-09-214-278-3
27	40.5	34.0	1238	4	US-09-214-278-5

```

28 40.5 34.0 1248 4 US-08-882-046-6 Sequence 6, Appli
29 40 33.6 196 2 US-08-933-750C-35 Sequence 35, Appl
30 40 33.6 196 4 US-09-234-613-35 Sequence 35, Appl
31 40 33.6 700 3 US-08-931-952-2 Sequence 2, Appli
32 40 33.6 700 3 US-08-272-247-2 Sequence 2, Appli
33 40 33.6 700 5 PCT-US95-08560-2 Sequence 2, Appli
34 40 33.6 839 4 US-09-197-636-2 Sequence 2, Appli
35 40 33.6 839 4 US-09-197-636-4 Sequence 2, Appli
36 40 33.6 839 4 US-09-197-636-8 Sequence 8, Appli
37 39 32.8 82 4 US-09-071-297-22 Sequence 2, Appli
38 39 32.8 384 3 US-09-311-170-2 Sequence 2, Appli
39 39 32.8 385 1 US-08-361-920-23 Sequence 23, Appl
40 39 32.8 385 1 US-08-479-939-23 Sequence 23, Appl
41 39 32.8 385 1 US-08-483-432-23 Sequence 23, Appl
42 39 32.8 445 5 PCT-US94-05387-7 Sequence 7, Appli
43 39 32.8 895 1 US-08-123-161A-8 Sequence 8, Appli
44 39 32.8 895 1 US-08-483-278-8 Sequence 8, Appli
45 39 32.8 1091 3 US-08-633-768A-2 Sequence 2, Appli

```

#### ALIGNMENTS

```

RESULT 1
US-09-007-999-2
; Sequence 2, Application US/09007999
; Patent No. 6087559
; GENERAL INFORMATION:
; APPLICANT: NICHOLS, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0356D
; CURRENT APPLICATION NUMBER: US/09/007,999
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-007-999-2

```

```

Query Match          90.8%; Score 108; DB 3; Length 1475;
Best Local Similarity 85.7%; Pred. No. 8.2e-09;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ANNHVSIVEAWSNDTPYLHD 21
   ||:|:|:|:|:|:|:|:|:|
Db 481 ANDHLSILEAWSNDTPYLHD 501

```

```

RESULT 2
US-09-210-361-2
; Sequence 2, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: NICHOLS, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07

```

; EARLIER APPLICATION NUMBER: 09/008,172  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-210-361-2

Query Match 90.8%; Score 108; DB 4; Length 1475;  
Best Local Similarity 85.7%; Pred. No. 8.2e-09;  
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ANNHVSIVEAWSNDNDTPYLHD 21  
||:|:|:|:|:|:|:|:|:|  
481 ANDHLSILEAWSNDNDTPYLHD 501

RESULT 3  
US-09-210-361-4  
; Sequence 4, Application US/09210361  
; Patent No. 6284479  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starches and  
; FILE REFERENCE: 0357CR  
; CURRENT APPLICATION NUMBER: US/09/210,361  
; CURRENT FILING DATE: 1998-12-11  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/009,620  
; EARLIER FILING DATE: 1998-01-20  
; EARLIER APPLICATION NUMBER: 08/485,243  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/008,172  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 1375  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-210-361-4

Query Match 83.2%; Score 99; DB 4; Length 1375;  
Best Local Similarity 81.0%; Pred. No. 2e-07;  
Matches 17; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Oy 1 ANNHVSIVEAWSNDNDTPYLHD 21  
||:|:|:|:|:|:|:|:|:|  
Db 507 ANDHLSILEAWSNDNDTPYLHD 527

RESULT 4  
US-09-008-172-2  
; Sequence 2, Application US/09008172  
; Patent No. 6127602  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; FILE REFERENCE: 0358D  
; CURRENT APPLICATION NUMBER: US/09/008,172

; CURRENT FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-008-172-2

Query Match 62.2%; Score 74; DB 3; Length 1430;  
Best Local Similarity 66.7%; Pred. No. 0.0019;  
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Oy 1 ANNHVSIVEAWSNDNDTPYLHD 21  
||:|:|:|:|:|:|:|:|:|  
Db 495 AINHLSILEAWSNDNDPQYNKD 515

RESULT 5  
US-09-210-361-6  
; Sequence 6, Application US/09210361  
; Patent No. 6284479  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starches and  
; FILE REFERENCE: 0357CR  
; CURRENT APPLICATION NUMBER: US/09/210,361  
; CURRENT FILING DATE: 1998-12-11  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/009,620  
; EARLIER FILING DATE: 1998-01-20  
; EARLIER APPLICATION NUMBER: 08/485,243  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/008,172  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 6  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-210-361-6

Query Match 62.2%; Score 74; DB 4; Length 1430;  
Best Local Similarity 66.7%; Pred. No. 0.0019;  
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Oy 1 ANNHVSIVEAWSNDNDTPYLHD 21  
||:|:|:|:|:|:|:|:|:|  
Db 495 AINHLSILEAWSNDNDPQYNKD 515

RESULT 6  
US-08-793-824-2  
; Sequence 2, Application US/08793824  
; Patent No. 5981838  
; GENERAL INFORMATION:  
; APPLICANT: Simpson, Christine Lynn  
; APPLICANT: Giffard, Philip Morrison  
; APPLICANT: Jacques, Nicholas Anthony  
; TITLE OF INVENTION: Genetic Manipulation of Plants to  
; FILE REFERENCE: 0358D  
; CURRENT APPLICATION NUMBER: Increase Stored Carbohydrates  
; NUMBER OF SEQUENCES: 2

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Griffith Hack & Co  
;; STREET: Level 8, 168 Walker Street  
;; CITY: No. 5981838th Sydney  
;; STATE: New South Wales  
;; COUNTRY: Australia  
;; ZIP: 2060  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/793,824  
;; FILING DATE:  
;; CLASSIFICATION: 800  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: AU PM7643  
;; FILING DATE: 24-AUG-1994  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 61 2 9957 5944  
;; TELEFAX: 61 2 957 6288  
;; TELEX: 26547  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1577 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: not relevant  
;; MOLECULE TYPE: protein  
;; ORIGINAL SOURCE:  
;; ORGANISM: Streptococcus salivarius  
;; US-08-793-824-2

Query Match 47.9%; Score 57; DB 2; Length 1577;  
Best Local Similarity 61.1%; Pred. No. 1;  
Matches 11; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 4 HVSIWEAWSNDTPYLHD 21  
|:||||| | | | |  
Db 591 HLSLEAWSNDHQYKND 608

RESULT 7  
US-08-836-442-4  
; Sequence 4, Application US/08836442  
; Patent No. 5990293  
; GENERAL INFORMATION:  
; APPLICANT: DOCHERTY, Andrew, J.P.  
; APPLICANT: SLOCOMBE, Patrick, M.

;; TITLE OF INVENTION: A HUMAN METALLOPROTEINASE  
;; TITLE OF INVENTION: VARIANTS THEREOF AND DNA SEQUENCES CODING THEREFOR  
;; NUMBER OF SEQUENCES: 4  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP  
;; STREET: 130 Water Street  
;; CITY: Boston  
;; STATE: MA  
;; COUNTRY: USA  
;; ZIP: 02109  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: DOS  
;; SOFTWARE: FastSeq for Windows Version 2.0  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/836,442  
;; FILING DATE: 01-MAY-1997  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/GB96/02181  
;; FILING DATE: 13-MAR-1997

;; APPLICATION NUMBER: GB 9612150.4  
;; FILING DATE: 11-JUN-1996  
;; APPLICATION NUMBER: GB 9526229.1  
;; FILING DATE: 21-DEC-1995  
;; APPLICATION NUMBER: GB 9521498.7  
;; FILING DATE: 20-OCT-1995  
;; APPLICATION NUMBER: GB 95521495.3  
;; FILING DATE: 20-OCT-1995  
;; APPLICATION NUMBER: GB 9518023.8  
;; FILING DATE: 05-SEP-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Resnick, David S  
;; REGISTRATION NUMBER: 34,235  
;; REFERENCE/DOCKET NUMBER: 47425  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 617-523-3400  
;; TELEFAX: 617-523-6440  
;; TELEX:  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 200 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: No. 5990293e  
;; US-08-836-442-4

Query Match 35.3%; Score 42; DB 2; Length 200;  
Best Local Similarity 50.0%; Pred. No. 22;  
Matches 8; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

Qy 2 NNHVSIV--EAWSDND 15  
| | | | | | | | | |  
Db 47 NTHVALVGMETWTKD 62

RESULT 8  
5290690-11  
; Patent No. 5290690  
; APPLICANT: MRABET, NADIR; LASTERS, IGNACE; STANSSENS, PATRICK  
; MATTHYSSENS, GASTON; WODAK, SHOSHANA; QUAX, WILHELMUS J.  
; TITLE OF INVENTION: METHODS AND MEANS FOR CONTROLLING THE  
; STABILITY OF PROTEINS  
; NUMBER OF SEQUENCES: 22  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/398,706  
; FILING DATE: 25-AUG-1989  
; SEQ ID NO: 11  
; LENGTH: 334  
5290690-11

Query Match 35.3%; Score 42; DB 6; Length 334;  
Best Local Similarity 53.8%; Pred. No. 40;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 8 VEAWSNDTPYLH 20  
| : | | | | | |  
Db 308 VVSWYDNETGYSH 320

RESULT 9  
US-09-362-473-6  
; Sequence 6, Application US/09362473  
; Patent No. 6218169  
; GENERAL INFORMATION:  
; APPLICANT: Cahoon, Edgar B.  
; APPLICANT: Cahoon, Rebecca E.  
; APPLICANT: Falco, S. Carl  
; APPLICANT: Morgante, Michele  
; APPLICANT: Rafalski, J. Antoni  
; APPLICANT: Hutz, William D.

; APPLICANT: Kinney, Anthony J.  
; TITLE OF INVENTION: Aromatic Amino Acid Catabolism Enzymes  
; FILE REFERENCE: BB-1197  
; CURRENT APPLICATION NUMBER: US/09/362.473  
; CURRENT FILING DATE: 1999-07-28  
; EARLIER APPLICATION NUMBER: 60/094.783  
; EARLIER FILING DATE: JULY 31, 1998  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 6  
; LENGTH: 455  
; TYPE: PRT  
; ORGANISM: Glycine max  
US-09-362-473-6

Query Match 35.3%; Score 42; DB 4; Length 455;  
Best Local Similarity 50.0%; Pred. No. 57;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
QY 8 VEANSNDNTPYLHD 21  
Db 270 VVAMHGNTVPYMYD 283

RESULT 10  
US-08-836-442-3  
; Sequence 3, Application US/08836442  
; Patent No. 5990293  
; GENERAL INFORMATION:  
; APPLICANT: DOCHERTY, Andrew, J.P.  
; APPLICANT: SLOCOMBE, Patrick, M.  
; TITLE OF INVENTION: A HUMAN METALLOPROTEINASE  
; TITLE OF INVENTION: VARIANTS THEREOF AND DNA SEQUENCES CODING THEREFOR  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP  
; STREET: 130 Water Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/836.442  
; FILING DATE: 01-MAY-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB96/02181  
; FILING DATE: 13-MAR-1997  
; APPLICATION NUMBER: GB 9612150.4  
; FILING DATE: 11-JUN-1996  
; APPLICATION NUMBER: GB 9526229.1  
; FILING DATE: 21-DEC-1995  
; APPLICATION NUMBER: GB 9521498.7  
; FILING DATE: 20-OCT-1995  
; APPLICATION NUMBER: GB 95521495.3  
; FILING DATE: 20-OCT-1995  
; APPLICATION NUMBER: GB 9518023.8  
; FILING DATE: 05-SEP-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Resnick, David S  
; REGISTRATION NUMBER: 34, 235  
; REFERENCE/DOCKET NUMBER: 47425  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-523-3400  
; TELEFAX: 617-523-6440  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 529 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 5990293e  
US-08-836-442-3

Query Match 35.3%; Score 42; DB 2; Length 529;  
Best Local Similarity 50.0%; Pred. No. 68;  
Matches 8; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 2 NNHYSIV--EAWSOND 15  
Db 246 NTHVALGMEIWDKD 261

RESULT 11  
US-07-665-792E-9  
; Sequence 9, Application US/07665792E  
; Patent No. 5281694  
; GENERAL INFORMATION:  
; APPLICANT: Baseman, Joel B.  
; APPLICANT: Su, C. J.  
; APPLICANT: Dallo, S. F.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR  
; TITLE OF INVENTION: PRODUCTION OF MYOPLASMAL ADHESINS  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ARNOLD, WHITE & DURKEE  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: U.S.A.  
; ZIP: 77210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5-1/4" DSD Diskette  
; COMPUTER: Compaq (IBM PC Compatible)  
; OPERATING SYSTEM: MS DOS 3.31  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/665.792E  
; FILING DATE: 19910607  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/118,967  
; FILING DATE: No. 5281694. 10, 1987  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/588,886  
; FILING DATE: July 27, 1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/616,111  
; FILING DATE: No. 5281694. 21, 1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/697,349  
; FILING DATE: May 8, 1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Melinda Patterson  
; REGISTRATION NUMBER: 33,062  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (713) 787-1592  
; TELEFAX: (713) 789-2679  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1627 Amino Acids  
; TYPE: AMINO ACID  
; TOPOLOGY: Linear  
; MOLECULE TYPE: Genomic DNA  
; DESCRIPTION: Deduced amino acid sequence of p1  
; DESCRIPTION: protein  
; HYPOTHETICAL: No  
; ORIGINAL SOURCE:

ORGANISM: Mycoplasma pneumoniae  
STRAIN: M129-B16  
IMMEDIATE SOURCE:  
CLONE: Gene cloned in pUC19 in E. coli HB101, ATCC  
CLONE: Accession Number 67560  
FEATURE:  
NAME/KEY: Amino acid sequence of p1 protein  
LOCATION: Amino Acid Numbers: 1 to 1627  
US-07-665-792E-9

Query Match 35.3%; Score 42; DB 1; Length 1627;  
Best Local Similarity 39.3%; Pred. NO. 2.5e+02;  
Matches 11; Conservative 2; Mismatches 3; Indels 12; Gaps 1;

QY 5 VSIVEA-----WSDNDTPYLH 20  
DB 1214 VSVVEASAYKPTSSGQQTSSPYLH 1241

ULT 12  
08-750-717-2  
Sequence 2, Application US/08750717  
Patent No. 6180109

GENERAL INFORMATION:  
APPLICANT: MOORMANN, Robertus J. M.  
TITLE OF INVENTION: Nucleotide Sequences of Pestivirus  
TITLE OF INVENTION: Strains, Polypeptides Encoded by These Sequences and Use  
TITLE OF INVENTION: Thereof for Diagnosis and Prevention of Pestivirus  
TITLE OF INVENTION: Infections  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: YOUNG & THOMPSON  
STREET: 745 South 23rd Street  
CITY: Arlington  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22202

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/750,717  
FILING DATE: 24-DEC-1996  
CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 94201743.5  
FILING DATE: 17-JUN-1994  
APPLICATION NUMBER: WO PCT/NL95/00214  
FILING DATE: 16-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: PATCH, Andrew J.  
REGISTRATION NUMBER: 32,925  
REFERENCE/DOCKET NUMBER: BO 39123  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-521-2297  
TELEFAX: 703-685-0573  
TELEX: 248425 EMBON

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3898 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-750-717-2

Query Match 35.3%; Score 42; DB 4; Length 3898;  
Best Local Similarity 41.2%; Pred. NO. 6.7e+02;

Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;  
QY 3 NHVSIVEAWSNDTPYL 19  
DB 3678 SHTPIQVRWSDNTSSYM 3694

RESULT 13  
US-08-277-231A-3  
Sequence 3, Application US/08277231A  
Patent No. 5643725

GENERAL INFORMATION:  
APPLICANT: Green, Bruce A.  
APPLICANT: Brinton, Charles C.  
TITLE OF INVENTION: Sequence and Analysis of LKP Pillin  
Patent No. 5643725  
TITLE OF INVENTION: Structural Genes and The LKP Pilli Operon of No. 5643725type  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02173

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/277,231A  
FILING DATE: 19-JUL-1994  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Alice O.  
REGISTRATION NUMBER: 32,542  
REFERENCE/DOCKET NUMBER: ACC94-02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 259 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-277-231A-3

Query Match 34.5%; Score 41; DB 1; Length 259;  
Best Local Similarity 37.5%; Pred. No. 43;  
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 2 NNHVSIVEAWSNDTP 17  
DB 72 DDSAALVQAWIDGNP 87

RESULT 14  
US-08-473-750-6  
Sequence 6, Application US/08473750  
Patent No. 5834187

GENERAL INFORMATION:  
APPLICANT: Green, Bruce A.  
APPLICANT: Brinton, Jr., Charles C.  
TITLE OF INVENTION: Sequence and Analysis of LKP Pillin  
Patent No. 5834187  
TITLE OF INVENTION: Structural Gene and the LKP Pilli Operon of No. 5834187 5786  
TITLE OF INVENTION: Haemophilus Influenzae

; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: US  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/473,750  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/277,321  
; FILING DATE: 19-JUL-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Carroll, Alice O.  
; REGISTRATION NUMBER: 33,542  
; REFERENCE/DOCKET NUMBER: ACC94-02B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-861-6240  
; TELEFAX: 617-861-9540  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 259 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-473-750-6

Query Match 34.5%; Score 41; DB 2; Length 259;  
Best Local Similarity 37.5%; Pred. No. 43;  
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 2 NNHVSIVEAWSNDTP 17  
:::|||||  
Db 72 DDSAALVQAWIDNGNP 87

## RESULT 15

US-08-477-326-6  
Sequence 6, Application US/08477326  
Patent No. 5968769

; GENERAL INFORMATION:  
; APPLICANT: Green, Bruce A.  
; APPLICANT: Brinton, Jr., Charles C.  
; TITLE OF INVENTION: Sequence and Analysis of LKP Pili  
; Patent No. 5968769  
; TITLE OF INVENTION: Structural Gene and the LKP Pili Operon of No. 5968769typable  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: US  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/477,326  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION:

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/277,231  
; FILING DATE: July 19, 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Carroll, Alice O.  
; REGISTRATION NUMBER: 33,542  
; REFERENCE/DOCKET NUMBER: ACC94-02A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-861-6240  
; TELEFAX: 617-861-9540  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 259 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-477-326-6

Query Match 34.5%; Score 41; DB 2; Length 259;  
Best Local Similarity 37.5%; Pred. No. 43;  
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 2 NNHVSIVEAWSNDTP 17  
:::|||||  
Db 72 DDSAALVQAWIDNGNP 87

Search completed: March 27, 2002, 13:59:30  
Job time: 583 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:20:25 ; Search time 1139.61 seconds  
(without alignments)  
5.116 Million cell updates/sec

Title: US-09-290-049a-12  
Perfect score: 119  
Sequence: 1 ANNHVSIVEAWSNDTPYLHD 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 3148936 seqs, 277657034 residues

1 number of hits satisfying chosen parameters: 3148936

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:\*

- 1: /cgn2\_6/ptodata/2/paa/PTCUS\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/paa/US080\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pep.\*
- 7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pep.\*
- 8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep.\*
- 9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pep.\*
- 10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep.\*
- 11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pep.\*
- 12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pep.\*
- 13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep.\*
- 14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep.\*
- 15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pep.\*
- 16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pep.\*
- 17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pep.\*
- 18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pep.\*
- 19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pep.\*
- 20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pep.\*
- 21: /cgn2\_6/ptodata/2/paa/US097\_COMB.pep.\*
- 22: /cgn2\_6/ptodata/2/paa/US098\_COMB.pep.\*
- 23: /cgn2\_6/ptodata/2/paa/US099\_COMB.pep.\*
- 24: /cgn2\_6/ptodata/2/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119	100.0	21	16	US-09-290-049-12
2	119	100.0	21	16	US-09-290-049-14
3	108	90.8	21	16	US-09-290-049-1
4	108	90.8	21	16	US-09-290-049-10
5	108	90.8	1475	19	US-09-557-848-2
6	108	90.8	1475	21	US-09-740-274-2
7	99	83.2	1375	21	US-09-740-274-4
8	74	62.2	21	16	US-09-290-049-11
9	74	62.2	1430	20	US-09-649-885-2
					Sequence 12, Appl
					Sequence 14, Appl
					Sequence 1, Appl
					Sequence 10, Appl
					Sequence 2, Appl
					Sequence 2, Appl
					Sequence 4, Appl
					Sequence 11, Appl
					Sequence 2, Appl

10	74	62.2	1430	21	US-09-740-274-6	Sequence 6, Appli
11	62	52.1	21	16	US-09-290-049-13	Sequence 13, Appl
12	55	46.2	2057	18	US-09-499-203-2	Sequence 2, Appli
13	53	44.5	196	19	US-09-595-298A-765	Sequence 765, App
14	50	42.0	1252	24	US-60-191-637-4899	Sequence 4899, Ap
15	48.5	40.8	401	1	PCT-US01-01309-222	Sequence 222, App
16	47	39.5	749	19	US-09-562-737-96	Sequence 96, Appl
17	46.5	39.1	220	24	US-60-212-413-164	Sequence 164, Appl
18	46.5	39.1	220	24	US-60-229-518-331	Sequence 331, App
19	46.5	39.1	330	24	US-60-207-422-75	Sequence 75, Appl
20	46.5	39.1	539	1	PCT-US01-01332-762	Sequence 762, App
21	46.5	39.1	539	21	US-09-764-875-762	Sequence 764, App
22	46	38.7	47	21	US-09-758-470-704	Sequence 31546, A
23	46	38.7	164	18	US-09-137-970-2	Sequence 2, Appli
24	46	38.7	880	15	US-09-137-970-2	Sequence 12250, A
25	46	38.7	880	24	US-60-167-217-12250	Sequence 9936, Ap
26	46	38.7	880	24	US-60-173-464-9936	Sequence 12251, A
27	46	38.7	880	24	US-60-191-637-12251	Sequence 9608, Ap
28	46	38.7	880	24	US-60-191-681-9608	Sequence 4571, Ap
29	45	37.8	199	21	US-09-738-626-4571	Sequence 661, App
30	45	37.8	306	19	US-09-595-298A-661	Sequence 660, App
31	45	37.8	340	19	US-09-595-298A-660	Sequence 1370, Ap
32	45	37.8	1270	24	US-60-242-679-1370	Sequence 22573, A
33	44	37.0	73	16	US-09-248-796-22573	Sequence 21844, A
34	44	37.0	129	16	US-09-248-796-21844	Sequence 13119, A
35	44	37.0	261	23	US-09-902-540-13119	Sequence 20736, A
36	44	37.0	391	24	US-60-324-109-20736	Sequence 4985, Ap
37	44	37.0	659	24	US-60-167-217-4985	Sequence 11, Appl
38	43.5	36.6	202	10	US-08-633-719-11	Sequence 11, Appl
39	43.5	36.6	202	16	US-09-280-197-11	Sequence 9810, Ap
40	43.5	36.6	656	23	US-09-902-540-9810	Sequence 3, Appli
41	43.5	36.6	1092	12	US-08-836-156-3	Sequence 3, Appli
42	43.5	36.6	1092	16	US-09-275-608-3	Sequence 5, Appli
43	43.5	36.6	1092	18	US-09-423-126-5	Sequence 891, App
44	43	36.1	40	24	US-60-196-190-891	Sequence 324, App
45	43	36.1	67	20	US-09-684-524-324	

ALIGNMENTS

RESULT 1  
US-09-290-049-12  
; Sequence 12, Application US/092900049  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: FDC98-01p2A  
; CURRENT APPLICATION NUMBER: US/09/290,049  
; CURRENT FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081,550  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/115,142  
; EARLIER FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 12  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: S. downei  
US-09-290-049-12

Query Match 100.0%; Score 119; DB 16; Length 21;  
Best Local Similarity 100.0%; Pred. No. 9.6e-12;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ANNHVSIVEAWSNDTPYLHD 21  
Db 1 ANNHVSIVEAWSNDTPYLHD 21

```

RESULT      2
US-09-290-049-14
; Sequence 14, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINE
; TITLE OF INVENTION: CARIES
; FILE REFERENCE: FDC98-0192A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. sobrinus
US-09-290-049-14

```

Query Match	100.0%	Score 119;	DB 16;
Best Local Similarity	100.0%	Pred. No. 9.6e-12;	Length 21;
Matches 21;	Conservative 0;	Mismatches 0;	Indels 0;
Gaps 0;			

QY 1 ANNHVSIVEAWSNDTPLYLHD 21  
| | | | | | | | | | | |  
Db 1 ANNHVSIVEAWSNDTPLYLHD 21

```

3
US-09-290-049-1
; Sequence 1, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CARIES
; FILE REFERENCE: FDC98-01p2A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: EAW peptide
US-09-290-049-1

```

Query Match 90.8%; Score 108; DB 16; Length 21;  
Best Local Similarity 85.7%; Pred. No. 5.7e-10;  
Matches 18; Conservative 3; Mismatches 0; Indels

```
QY 1 ANNHVSIVEAWSNDNDTPYLHD 21
    ||:||:||:||:||:|||||
DB 1 ANDHLSILEAWSNDNDTPYLHD 21
```

RESULT 4  
US-09-290-049-10  
; Sequence 10, Application US/09290049  
; GENERAL INFORMATION:

```

; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CRIES
; FILE REFERENCE: FDC98-01p2A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. mutans
; US-09-290-049-10

```

Query Match	90.8%	Score 108;	DB 16;	Length 21;
Best Local Similarity	85.7%;	Pred. No. 5.7e-10;		
Matches 18;	Conservative	3;	Mismatches 0;	Indels 0;
				Gaps 0;

```
Qy 1 ANNHVSIVEAWSNDTPYLHD 21
    ||:||:||:||:|||||
Db 1 ANDHLSILEAWSNDTPYLHD 21
```

```

RESULT 5
US-09-557-848-2
; Sequence 2, Application US/09557848
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0356D2
; CURRENT APPLICATION NUMBER: US/09/557,848
; CURRENT FILING DATE: 2000-04-26
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-557-848-2

```

Query Match	90.8%	Score 108;	DB 19;	Length 1475;
Best Local Similarity	85.7%	Pred. No. 7.4e+08;		
Matches 18; Conservative	3;	Mismatches 0;	Indels 0;	

```
Qy 1 ANNHVSIVEAWSNDTPYLHD 21
    ||:|:|:|:|:|:|:|:|:|:|
Db 481 ANDHLSILEAWSNDTPYLHD 501
```

RESULT# 6  
US-09-740-274-2  
; Sequence 2, Application US/09740274  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999



